A mathematical framework for human pluripotent stem cell behaviours

Laura Emily Wadkin

Thesis submitted for the degree of
Doctor of Philosophy

School of Mathematics, Statistics & Physics
Newcastle University
Newcastle upon Tyne
United Kingdom

September 2020

Acknowledgements

I would like to thank my supervisors Dr Nick Parker and Prof Anvar Shukurov for their guidance and support throughout this project. I am grateful to Dr Sirio Orozco-Fuentes for her essential advice and contributions and to Lesley Elliot for an introduction to the mathematics of stem cells.

This work would not have been possible without the biologists who dedicated their time, experiments and knowledge. Special thanks to Dr Irina Neganova and Prof Majlinda Lako, for their supervision, experimental facilities and biological expertise. Thanks to Dr Alex Laude for conducting the experimental image analysis, and Dr Valeria Chichagova and Dr Sanja Bojic for carrying out crucial experiments.

Further thanks to Georgina Swan, Dr Andrew Baggaley and Prof Rafael Barrio for their valuable contributions as co-authors.

Abstract

Human pluripotent stem cells, hPSCs, have unparalleled potential for developments in regenerative medicine, personalised medicine and drug discovery. For these promising clinical applications to become a reality, a deeper understanding of their complex behaviours across a multi-scale hierarchy is required. The use of techniques from mathematics and physics allows the identification of the systematic and universal behaviours inherent in a biological system, provides a framework for comparisons, tests and predictions, and can ultimately guide experimental decision making. We take a methodical approach to developing coherent models of hPSC behaviours, considering some of their key properties in isolation. The range of models developed includes a descriptive analysis of cell movements that leads to their association into pairs and further into colonies, the growth of clonal cell groups within a colony and the dynamics of intra-cellular pluripotency.

We consider the kinematics of single and pairs of hPSCs in two-dimensions using time-lapse microscopy imaging, quantifying their movements within a random walk framework and characterising their inherent correlation properties. This analysis reveals single cells perform an unusual anisotropic random walk along a local axis, with increased migration speeds in the direction of cell elongation. Pairs of cells in close proximity show a preference for moving in the same direction. The addition of a common biological marker (CellTracer™) negatively impacts the motility of both single and pairs of cells.

Clonal (genetically identical) hPSC colonies are required for many *in-vitro* applications. We consider the impact of spatial colony growth on undesirable clonality loss. Our experimental data show that colony populations are multi-modal, with a growth rate dependent on the number of founding cells. From this data, we extract the parameters for a stochastic exponential growth model which can be used to predict the time at which clonality is lost due to the merging of neighbouring colonies at different seeding densities.

Finally, we examine the internal regulation of cell pluripotency – the defining characteristic of hPSCs which allows for their differentiation into other cell types. Pluripotency is regulated by a complex network of pluripotency transcription factors (PTFs). We use experimental data to quantify the temporal regulation of the PTF OCT4 in a growing stem cell colony using the Hurst exponent, autocorrelation and diffusion analysis. We then present a 'tool kit' of temporal models which can be used to capture the fluctuation of PTFs (as a proxy for cell pluripotency) and evaluate the successes and limitations of each model.

Throughout this work the mathematics is rigorously underpinned by experimental results. Our global aim is to apply a variety of mathematical tools to deepen our understanding of stem cell behaviours and bridge the gap between experiments and theory.

Contents

Chapter 1

Stem cells and mathematics

Stem cells are at the forefront of modern biological research. This thesis combines experimental analysis and mathematical modelling to deepen our understanding of some of the key behaviours of stem cells. In this chapter we introduce stem cells and their biological properties, along with their promise for developments in drug discovery and regenerative medicine, and the current challenges in efficient laboratory culturing. Mathematics is a powerful tool to bridge the gap between experiments and theory and advance stem cell research. We discuss the advantages of an inter-disciplinary approach and review mathematical techniques for modelling stem cells.

1.1 Stem cells and their significance

Stem cells are undifferentiated cells, capable of differentiating into the various types of specialised cells that comprise the tissues of a multi-cellular organism [1]. It is this, along with their ability to self-renew through repeated divisions, producing two genetically identical daughter cells each time, that drives embryonic development and allows them to act as self-regulatory repair systems in some tissues, replacing other cells when they are dead or damaged [2–4].

Potency is a measure of a cell's potential to differentiate into a specialised cell. The highest level of potency, *totipotency*, is the ability to differentiate into any cell type, including placental cells. Human stem cells which are *pluripotent* (one level down from totipotent), are known as *human pluripotent stem cells* (hPSCs) and have the ability to differentiate into any cell type in the adult human body. Pluripotent stem cells are more readily available than totipotent cells, making them ideal for experiments and scientific research [5].

Stem cells can be artificially induced to form certain specialised cells types through targeted differentiation; for example, pancreatic cells, cardiomyocytes (heart muscle cells) and osteoblasts (a cell from which bone develops) have all been successfully induced from hPSCs [6–8]. It is for this reason that hPSCs are at the forefront of modern biological research and hold great promise for developments in regenerative medicine [5, 9, 10]. Recently developed medical treatments use stem cells that have been forced to differentiate *in-vitro* into the specific cell type required to build replacement tissues for transplants. They can also be used as a direct treatment in ground-breaking stem cell therapies, in which an injection of stem cells encourages surrounding injured tissues to self-repair [11, 12]. The demand for stem cells in the pharmaceutical industry is growing as they provide

a highly controllable platform to study the causes of disease alongside the development and trials of drugs and genetic treatments [13–16].

hPSCs occur either as human embryonic stem cells (hESCs) derived from the early embryo, or human induced pluripotent stem cells (hiPSCs) which are derived by the genetic reprogramming of differentiated cells [17]. hiPSCs can be made from other readily available specialised cell tissue, like fibroblasts, by editing genes to return them to a pluripotent state – a reversal of the differentiation process. This approach, which received the 2012 Nobel Prize in Physiology or Medicine for its discovery, offers patient-specific hPSCs reducing the risk of transplanted tissue rejection [18] and avoids the ethical issues associated with hESCs [19].

In the lab, hPSCs are grown in mono-layer colonies of up to thousands of cells (Figure 1.1) from which they can be directed for specific experiments or therapies, or expanded to produce further hPSC colonies. These colonies are often founded from the seeding of a single cell onto a substrate which then divides to create a genetically identical population. New biomedical technologies require the efficient, large-scale production of hPSCs if the demands of the emerging industry are to be met [20]. Furthermore, applications of hPSCs in the clinic require tight control over the pluripotency, *clonality* (the proportion of identical cells that share a common ancestry) and differentiation trajectories *in-vitro*.

However, the existing procedures for large-scale experiments remain inefficient and expensive due to low cloning efficiencies of 1% to 27% (the percentage of single cells seeded that form a clone) [21, 22]. Understanding the factors which promote the efficient generation and satisfactory control of hPSC colonies (and their derivatives) is a key challenge. In Section 1.2 we set out the key biological properties of hPSCs across a complex multi-scale hierarchy and in Section 1.3 we discuss how mathematics is facilitating stem cell research.

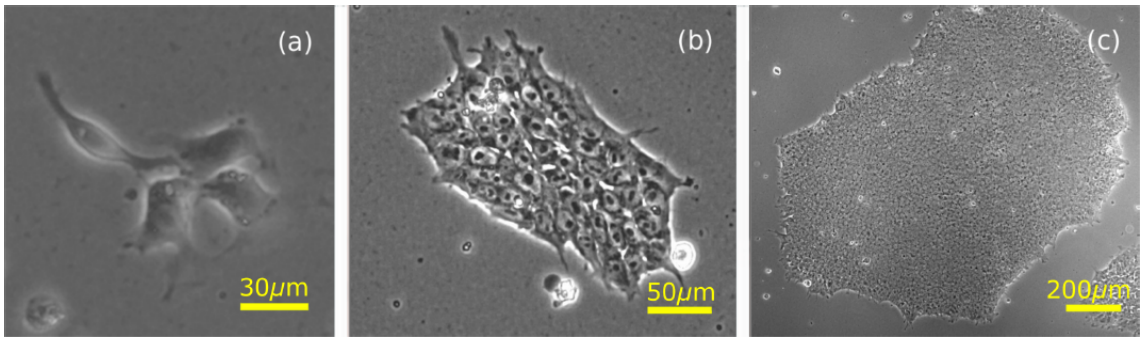

Figure 1.1: Microscopy images of hPSCs (hESCs) showing growing colonies from (a) a few cells up to colonies of (b) hundreds and (c) thousands.

1.2 Key biological properties of hPSCs

The satisfactory understanding and control of hPSC evolution remains elusive due to their complex behaviour over multiple scales: the intra-cellular scale (processes happening within cells), the cellular or micro-environment scale (the environmental effects on individual cells) and the colony scale (collective cell behaviours throughout colonies), as illustrated in Figure 1.2. Advances in imaging and molecular profiling (classification based on gene expression) have identified the core processes within the evolving colony [21, 23–25]. Here we outline these key biological properties of hPSCs.

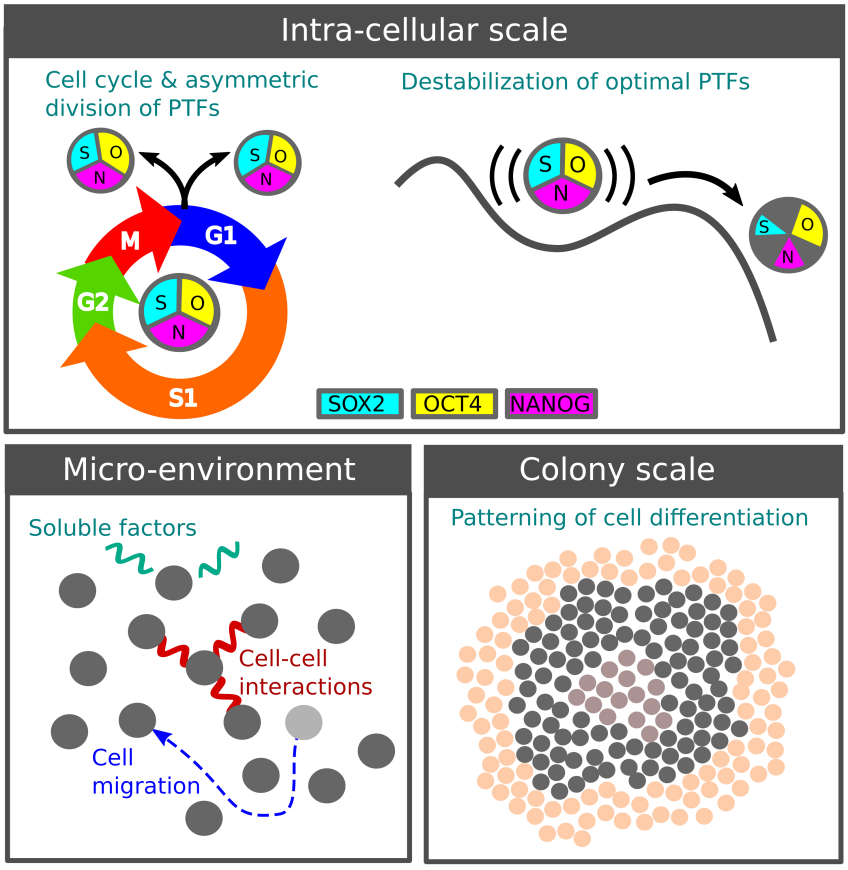

Figure 1.2: Scales of hPSC behaviour, the intra-cellular scale (e.g., cell cycle, division, inheritance of pluripotency transcription factors), the cell micro-environment (e.g., interaction with other cells, the medium and substrate) and colony-scale phenomena (e.g., patterning of differentiated cells).

1.2.1 Intra-cellular scale

The first key intra-cellular behaviour integral to hPSCs is the cell cycle. All cells (specialised and unspecialised) go through the cell cycle: the timed series of events controlling DNA replication and resulting in a cell division and two genetically identical daughter cells. The phases of the cell cycle are: G1 (growth phase), S (synthesis phase in which DNA is replicated), G2 (further growth) and M (mitosis, the cell division). The G1 phase is shortened for hPSCs, leading to more rapid proliferation than for somatic cells [26]. Cells in different stages of the cycle may exhibit different behaviours and it has been shown that cell fate decisions (whether cells live or die) are closely associated with the cell cycle [27, 28]. The division time of a cell is the time it takes to complete one full iteration of the cell cycle and is variable depending on cell lines and culture conditions. Microscopy images of a cell dividing into two daughter cells are shown in Figure 1.3. More commonly referred to by biologists is the doubling time of cells, the average time it takes for the population of cells to double. The doubling time of stem cells varies and can be affected by various environmental and chemical factors, including cell density and the colony maturity [21, 29–32], but for hPSCs it is generally accepted to be between 12 and 24 hours [33].

Every stem cell ultimately either remains pluripotent, differentiates into a specialised cell, or dies. This is known as the cell fate decision. There are two major types of cell death, apoptosis, irreversible regulated programmed death, and necrosis, death triggered by disease or trauma to the cell. Here we consider cell death for hPSCs generally, with no distinction between the two types. High rates of spontaneous death make hPSCs difficult to maintain in culture and the pathways that lead to such high rates of cell death is a continuing area of research [31, 34–38]. Interestingly, daughter cells resulting from a division show continued patterns of cell death, a major bottleneck restricting colony growth [21].

The second key intra-cellular behaviour of hPSCs is pluripotency regulation. The maintenance of pluripotency depends on the stable inter-regulation of pluripotency transcription factors (PTFs) [39], mainly by the genes OCT4, SOX2 and NANOG [40].

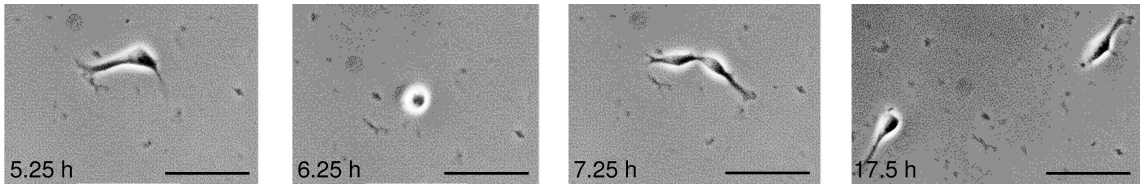

Figure 1.3: Microscopy images of a single hESC undergoing mitosis and dividing into two genetically identical daughter cells. The scale bar represents 50 μm. The images are from Experiment 1 which we discuss in Chapters 2 and 3.

We discuss the dynamics of PTFs in more detail in Section 1.3.2 and Chapter 5. Fluctuations of the PTF abundances are believed to cause the variation in pluripotency in different sub-populations [39]. Moreover, recent work suggests that the PTFs are inherited asymmetrically as a cell divides [41], biasing the fate of daughter cells and contributing to colony heterogeneity. Destabilisation and the interaction of these PTFs with chemical signalling pathways triggers differentiation, the departure from the pluripotent state [39, 42] towards specific cell fates (differentiation) [43]. The cell cycle also affects pluripotency and vice versa [40, 44, 45]. Although there are protocols for reducing the rates of undesirable spontaneous differentiation [46, 47] it remains a problem for the large-scale culturing of pluripotent colonies.

1.2.2 Micro-environment

The local environment of the cell is key to its *in-vitro* evolution. One of the leading environmental factors affecting hPSCs is the substrate on which they are grown. Substrates may either consist of a layer of mouse or human 'feeder' cells or a protein substrate, with the latter growing in popularity for clinical applications since they avoid the risk of genetic contamination. The substrate influences pluripotency [48] and mobility [49] through its growth factors and adhesion forces. There are protocols for inducing directed differentiation to generate desired specialised cell types [50] and it has been suggested that the local micro-environment modulates the intrinsic parameters that can be used to influence differentiation trajectories [51].

As well as the substrate, cell movements and cell-cell interactions are also important. hPSCs benefit from being in colonies where they exhibit higher viability and pluripotency [52]. Their migration is achieved through adaptations in cell morphology via the reorganisation of the actin cytoskeleton to form a leading edge pseudopodia [53]. We discuss the mechanics of individual cell migration in greater detail in Chapters 2 and 3. Unregulated cell migration *in-vitro* can cause clonality loss as the cell population grows, undesirable when a genetically identical clonal population is required [54–56]. Furthermore, anomalous cell migration has been linked to deviations in the undifferentiated state of hiPSCs [57]. A thorough understanding of the migration of hESCs is needed to optimise *in-vitro* clonality and facilitate the development of therapies for migration related disorders.

There are also mechanical effects of the increased culture density as colonies grow. As colonies become denser, the mutual mechanical pressure of the hPSCs can affect the cell cycle (prolonging the G1 phase), reduce colony proliferation and increase the cell propensity for differentiation [58].

1.2.3 Colony scale

Perhaps most intriguing, yet least understood, are behaviours that emerge on a colony scale. The promotion of pluripotency in larger colonies shows that single cells are influenced by the whole colony [51, 59]. Indeed, it has been suggested that pluripotency is a collective statistical property of cells [60], rather than a well-defined property of individual cells.

Further colony-scale effects are evident in the spatial patterning of the cell fates after differentiation. Mechanical forces and chemical signals operating over distances larger than the cell separation influence single-cell genetic expression to form bands of differentiated cells [61–63] (illustrated in Figure 1.2 and shown experimentally in Figure 1.4); these structures are enhanced under imposed boundaries, emphasising the role of mechanical forces [63, 64]. With further understanding, mechanical effects and boundaries could be harnessed to engineer specific desired differentiated cells [65].

Image analysis also provides evidence that the cell cycle is similarly spatially ordered within colonies [62]. The cells directly on the colony border are more likely to be in the G1 phase, and cells one to two layers into the colony (one to two cells from the colony edge) show increased proportions of cells in the G2 phase when compared to the with-in colony cells [62].

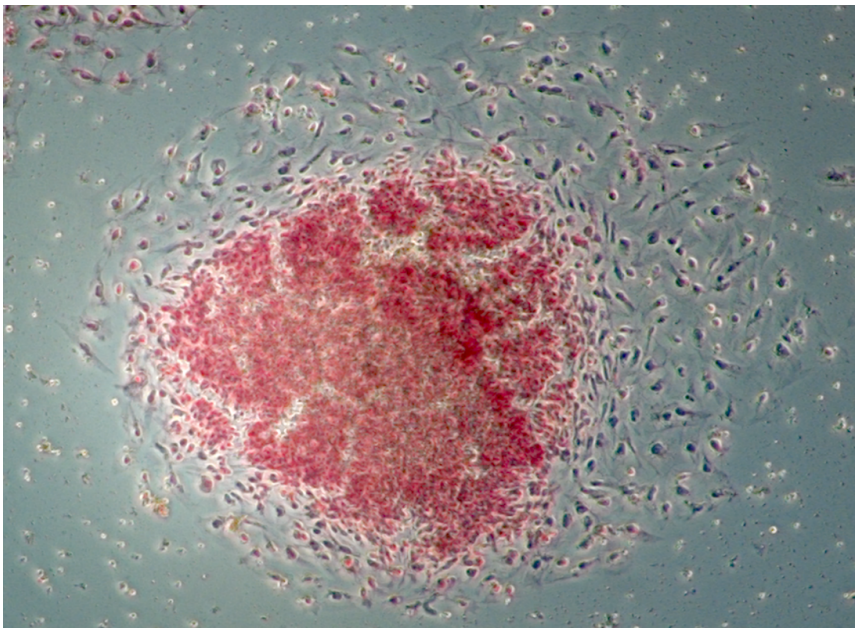

Figure 1.4: A colony of hiPSCs induced from fibroblast cells which have been stained to show their pluripotency; the most pluripotent cells are bright red. The band of darker cells around the edge of the colony are cells that have differentiated. The microscopy image was taken under $\times 10$ magnification [66].

hPSC colonies also exhibit collective migration, both *in-vitro* and in the human body [67, 68]. This coordinated migration of large numbers of hPSCs is essential in tissue generation [69] and wound healing [70]. Upon differentiation, the mass accumulation (growth rate of colony) of cells does not change radically, but coordinated cell movement is greatly reduced [71].

It is this broad range of complex behaviours across multiple scales which makes the improved understanding and control of both hPSCs and hiPSCs challenging. However, as we present in the next section, the quantitative analysis and mathematical modelling of these behaviours can help to untangle and characterise some of their systematic properties and guide experimental developments.

1.3 How can mathematics help?

A shorter version of this section has been submitted as a chapter ("An introduction to the mathematical modelling of iPSCs") for the book "Recent advances in iPSC technology" [72].

Mathematics is a powerful tool to achieve a deeper understanding of biological systems [73–75]. The application of mathematics to biology has led to many significant achievements in medicine and epidemiology (predicting the spread of 'mad cow' disease [76, 77] and influenza [78]), evolutionary biology [79] and cellular biology (descriptions of chemotaxis [80] and predicting cancer tumour growth [81]). Similarly, the use of mathematics in stem cell research is advancing current knowledge of the underlying behaviours discussed in Section 1.2 which may be difficult to elucidate experimentally and guiding experimental optimisations and protocol development [82–85]. An excellent introduction to the general subject of mathematical biology can be found in [74] and [75].

Mathematical biology uses established concepts from mathematics and theoretical physics to reveal generic aspects of biological behaviour and identify their underlying causes. Mathematical and computational modelling provides a framework for rigorous characterisation, the prediction of observations, and a profound understanding of the underlying natural processes. The wide applicability of these notions to hPSCs, including hiPSCs, is evident due to their systematic behaviours. There are many biological properties for which mathematical concepts are pertinent: the idea of pluripotency as a potential and statistical macrostate [60], cell fates as 'steady states' [39, 86], random walks and diffusive migrations of pluripotent stem cells [87–89], and the emergence of Turing-like spatial patterning [90] are just a few relevant examples. The first mathematical model of stem cells in 1964 used stochastic techniques to capture cell

fate decisions [91] and has since been extended to cover gene regulation [92–94], molecular states and the cell cycle [95], and cell population dynamics [96] based on experimental results.

In particular, when mathematical models are rigorously underpinned and validated by experimental observations, the reciprocal benefit for experimentation can be profound: an example is the development of an experimentally trained model of mouse iPSC programming, which led to strategies for marked improvements in reprogramming efficiency [83]. There has been a recent focus in complex systems to ensure such models are designed with a specific purpose in mind, are fit for this purpose, and are guided by experimental data [97].

Arguably the most important obstacle in expanding applications of mathematical modelling of stem cell biology is the natural gap between the knowledge and research approaches of experimentalists and theoreticians. Bridging this gap would lead to significant advances, as physics has demonstrated so spectacularly. It is, however, easy to understand that the process of narrowing the gap is difficult and requires motivation and devotion on both sides. In this section we review common mathematical ideas and approaches for modelling stem cell behaviour at a level that does not require more than elementary mathematics; it cannot replace systematic discussion of mathematical biology [74, 75], but we hope it can help guide dialogue between experimentalists and mathematicians.

An interdisciplinary approach combining mathematics and experimentation not only furthers our understanding of the underlying biological system, but also has direct implications for experimental optimisation. Below we list some of the key advantages of this approach, framed by recent advances in stem cell biology.

1. **Identification of systematic behaviours**

   The use of concepts from mathematics allows the identification and verification of systematic behaviours within biological systems. For example, the mathematical analysis of hPSC migration experiments within a random walk framework (including the results we present in Chapters 2 and 3) has led to the quantification and classification of the systematic migration of single and pairs of cells [87–89].

2. **Identifying universal behaviours**

   Not only does mathematics reveal the systematic behaviours, it can also highlight those which are universal. These universal behaviours reflect fundamental features of a biological system and are captured by identical or closely related governing equations; in turn, these highlight identical or related biological behaviour. Such models are then easily adaptable with simple parameter changes to different cell lines and experimental conditions. Even simple and well-established mathematical

concepts (such as logistic growth or random walks) can be rich and flexible in their specific applications. An example is the universal concept of Brownian motion used to describe the random dispersion of particles of an impressively diverse nature atoms [98], animals searching for food [99], people in a crowd [100] and, closer to our present subject, stem cell migration [87, 88, 101]. As the starting point for modelling cell migration, it provides a comprehensive, flexible and well understood framework for quantifying ensembles of cells and identifying any deviations from completely random movements and their implications on migration-induced clonality loss [21, 87]. Useful and interesting mathematical extensions include geometric Brownian motion [102] which can provide the addition of a systematic drift term (e.g., due to chemotaxis) and fractional Brownian motion [103] which could capture the persistence of cell movements.

3. **Framework for comparison, testing and predicting**
   A quantitative framework provides a clear basis for the comparison of different experiments from which the similarities and differences between biological conditions can be probed with quantitative precision and rigour. Firstly, statistical analysis of experimental data allows the quantification of stem cell behaviour which then informs the development of models [87, 88, 104]. Even at this early stage in the model building process, such qualifications reveal interesting insights for experimental comparisons. For example, the quantitative comparison of the migration of individual hPSCs under different experimental conditions (presented in Chapter 2) shows that the addition of CellTrace™, a labelling dye commonly used to track cell generations, significantly reduces cell motility [88]. Once developed, coherent mathematical models provide a non-invasive prognostic modelling tool for rigorous testing and the prediction of experimental behaviour. For iPSCs, mathematical reprogramming models have led to the suggestion of two mechanisms for reprogramming [95] and identified different modes of reprogramming dynamics [105].

4. **Guiding experimental decision making**
   Mathematical models and quantitative frameworks can assist with the objective decision making required in the development of experimental protocols. A simple hPSC spatial colony growth model (Chapter 4) can result in an equation allowing the prediction of the time at which adjacent colonies will first merge due to cell proliferation [82]. This result can guide experimentalists to select their cell seeding densities to optimise the colony clonality. Mathematical modelling is more efficient when in a continual feedback loop with experimental results; model development and refinement are informed by experimental data and the modelling results advise the

focus and analysis of the experiments. Computational simulations of mathematical models are also cheap to run compared to laboratory experiments. They provide an efficient cost-effective *in-silico* mode of experimentation.

5. **Strategic approaches for complex systems**
   Developing coherent models of hPSCs and their colonies remains challenging due to their complex behaviours across multiple scales, illustrated in Figure 1.2. Building coherent models of multi-scale systems requires a careful strategic approach. Mathematical techniques allow for a bottom-up strategy; key cell properties are first considered in isolation using the simplest models (minimal models) to identify the building blocks of the complexity. This systematic approach leads to the development of a hierarchy of models suitable for a range of behaviours and applications. These models can then be combined with others and developed to the level of sophistication required to capture the whole system. An informative review discussing the successes of multi-scale models in stem cell biology from a bioengineering approach is given in [106].

Although the range of mathematical tools applicable to building the hierarchical model of stem cell behaviour is vast, there are some common fundamental techniques underpinning most current models. These equations and techniques form the basis of more complex mathematical models currently under development for hPSC behaviour and hiPSC reprogramming. These models can be compiled into software 'toolkits' which provide a platform for model exploration under bespoke conditions [107].

In each section below we explore a different mathematical technique used for the analysis and modelling of stem cell properties, in a manner accessible for a reader without a mathematical background. We consider migration models, differential equations and agent-based modelling. We also have a concise review of the recent advances in the stem cell modelling available in [85].

1.3.1 Cell migration as a random walk

Migration is an intrinsic property of many different cell types, including stem cells, and represents a distinct feature of cell behaviour. Although the migration of different cell types may differ in important details, there are fundamental similarities. We discuss the specific mechanics of stem cell migration in Chapter 2. To begin, the internal mechanics of cell movement can often be ignored to focus on quantifying the properties of migration since they are largely independent of the specific mechanisms of motility. This quantification of cell motion provides a framework for comparison to other cell types and to cells under different experimental conditions. As in many other applications,

mathematical approaches provide an opportunity to separate, and then explore, generic and specific aspects of a complex phenomenon.

The classic mathematical description of cell motion is that of a random walk, a sequence of steps in random directions and of random lengths that form the migratory trajectory. Different cell types and different environments are uniquely characterised by the details of the random selection of the direction and distance of movement for each step. Although it is known that cells interact with each other and their environment, the large number of such factors, even if they are deterministic, and the complexity of their interactions justify and render efficiency to a probabilistic approach to simulating cell movement [84, 108]. In the following section we describe the mathematical framework for random walks, including correlated and biased random walks. Further mathematical details on the use of random walk models in biology can be found in [74, 109, 110].

Description of random walks

Analysing a tortuous, apparently random trajectory of a cells movement within a random walk framework leads to the identification of parameters which can be used to quantitatively describe cell migration and assist in predictive modelling [87–89, 101, 111, 112].

Firstly, consider the migration of a cell on a flat surface (a 2D system). The simplest is an isotropic random walk (IRW) that has no preferred direction in the cell movement. In this type of model, a continuous motion is approximated by a sequence of steps of a certain length. The shorter the step, the more accurate the approximation. The direction of each step is arbitrary and independent of the earlier movements (this independence is known as the Markovian property). It is natural to expect that cell migration *in-vitro* is isotropic in the absence of large-scale gradients in the environment, and far away from any boundaries.

An idealisation involved in the IRW description is the assumption that a cell moves along a straight line for a short period of time τ, covering a distance l, and then changes its direction of migration at random, with each direction having equal probability. Despite its simplicity, the model has been shown to capture a wide range of natural phenomena and is especially relevant to time-lapse imaging of cell migration where frames are taken at discrete times. The time, τ, is an intrinsic property of the migrating cell and characterises the internal motility mechanisms and is unrelated to the frequency of the image recording.

Although each individual cell movement is caused by many influences, the outcome of many such movements develops universal properties sensitive only to a few features of the instantaneous cell behaviour [87–89]. In particular, if two cells start close to each other, their trajectories will unavoidably diverge and, after a large number of steps, their separations will grow with time t as $\sqrt{t}$. Similarly, the size of a region occupied by a

group of cells, each involved in an independent random walk, grows with time as $\sqrt{t}$. This behaviour is observed in many systems including the spread (diffusion) of smell in air, dye in a liquid, heat from a flame, and is known as the diffusive behaviour.

To quantify such movement, consider a migrating cell with a current position $\mathbf{x}(t)$ and a starting position $\mathbf{x}_0 = \mathbf{x}(0)$. In each time step the cell makes a movement in a direction chosen at random. The diffusive nature of this random walk can be quantified by considering the mean square displacement (MSD) of the cells migratory trajectory. The MSD is a measure of the deviation of a cell from its starting position over time given by $\text{MSD} = \overline{(\mathbf{x}(t) - \mathbf{x}_0)^2}$ where the overbar denotes the average taken over a large number of cells. It is a measure of the portion of space explored by the random walker. When the cells motion is diffusive, the MSD increases linearly with time, $\text{MSD} = 2Dt$, where $D \approx l^2/2\tau$ is the diffusivity, a single summary parameter that characterises an IRW completely at long time or space intervals and fully characterises the cell motility. Thus, the long-term motion of cells is controlled not by their speed $v = l/\tau$, or distance travelled over a short time interval, but by their combination $D = l^2/2\tau = vl/2$. Depending on the details of cell motion and environments, the MSD may increase slower than t (then the motion is called sub-diffusive) or faster than t (this is super-diffusive behaviour). The variation of the MSD with time for each diffusion type and their corresponding trajectories are shown in Figure 1.5. Note that sub-diffusion can occur due to a combination of displacement and waiting times, as commonly seen in animal movements [109, 113, 114]. Super-diffusive motion has been seen in hESCs and has negative effects on clonality [21].

IRWs have been used to successfully describe the motion of hPSCs out of dense colonies (e.g., soon after seeding) [87, 88], including our results presented in Chapter 2. There are many ways to simulate the random walk migration of cells. For computational convenience, the cell positions can be restricted to a discrete lattice or grid, with cells only able to move between fixed positions on the lattice. Some models only allow one cell to occupy any given position on the grid at any time. There are also models in which the cells can move in continuous space. Both types of models, illustrated in Figure 1.6, are used for individual cell modelling and more complex models of colony kinematics.

A further useful measure to quantify the overall shape of a cell's trajectory is the directionality (sometimes referred to as the straightness index). The directionality, Δ, of a cell is defined as the ratio between its displacement r (i.e., the shortest distance between the cells current position and its initial position), and its total distance traversed d: $\Delta = r/d$ with $0 \leq \Delta \leq 1$. If the cell moves along a straight path then $r = d$, and the directionality has its maximum value, $\Delta = 1$. If the cell follows a long and tortuous trajectory, then d is much larger than r, and the directionality is low, $\Delta \approx 0$. Thus, the directionality quantifies how tangled and convoluted the cells trajectory is. In an IRW, on average, we have $r = \sqrt{\text{MSD}} = \sqrt{2Dt}$ and $d = vt$, where v is the instantaneous cell

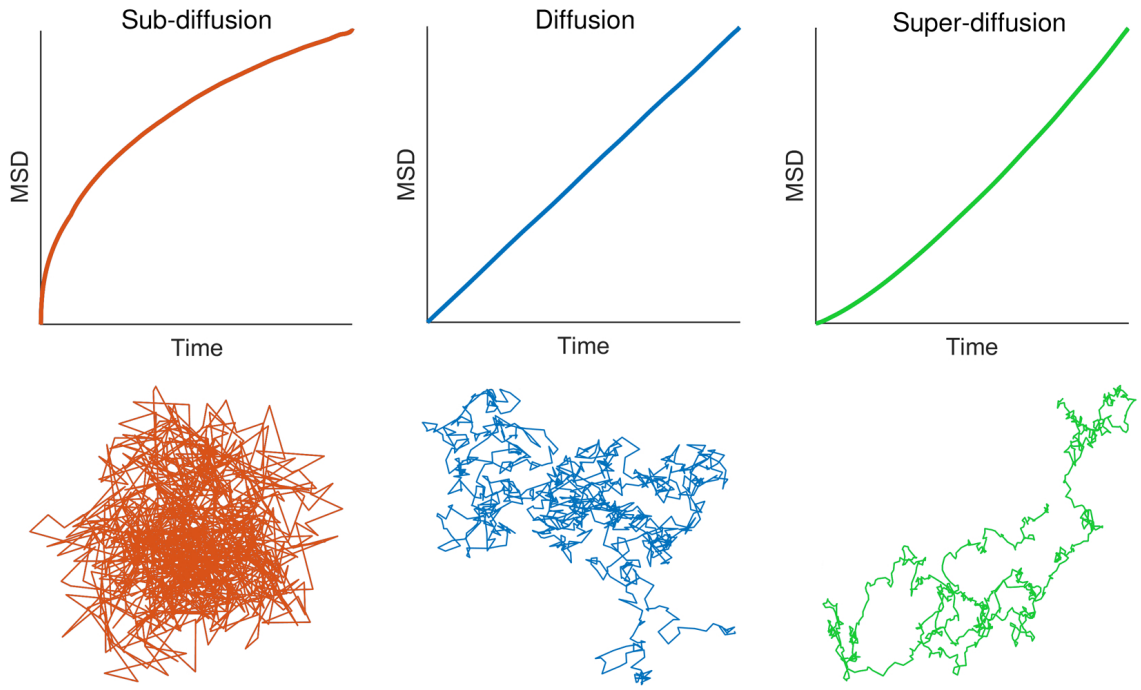

Figure 1.5: The MSD with time (top row) for sub-diffusive (left), diffusive (middle) super-diffusive (right) cell motion and examples of the corresponding cell trajectories (bottom row). Simulation details: all have the same number of steps of the same length. The postitions are updated as $\mathbf{x}_{i+1} = \mathbf{x}_i + \xi_i$, with ξ_i drawn from Brownian noise for the diffusive simulation, and fractional Brownian noise with Hurst exponents of 0.2 and 0.65 for sub- and super-diffusion, respectively.

speed and t is the time elapsed since the start of the motion. Thus, $\Delta = \sqrt{\tau/t}$, where τ is the duration of a cells persistent motion. This quantity is closely related to the tortuosity, similarly characterising the shape of convoluted trajectories commonly used in the analysis of animal movements [115, 116].

When the time scale τ of the IRW is much shorter than the duration t of an experiment, the discrete nature of the random walk model becomes insignificant. The resulting continuous random walk is described by the diffusion equation which can be considered in addition to, or instead of, simulations of individual random walkers. The diffusion equation is a partial differential equation for the density $n = n(x, y, t)$ of diffusing cells (the number of cells per unit area):

$$\frac{\partial n}{\partial t} = D\left(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2}\right). \tag{1.1}$$

The equation expresses the fact that the rate of change of the cell density over time $\partial n/\partial t$ is related to the diffusivity D and the variation of n in space (the coordinates x and y). Further details of how the equation is derived from random walk theory and its

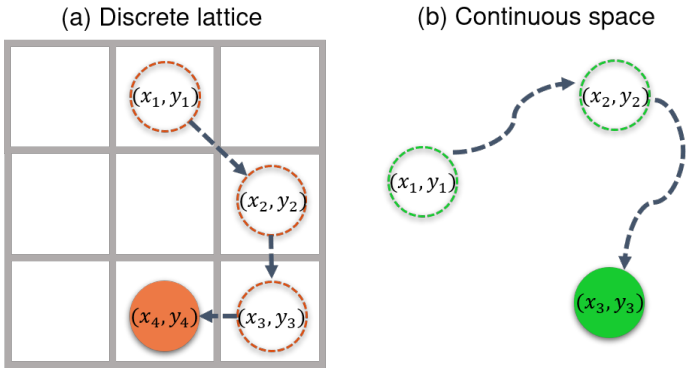

Figure 1.6: The random walk migration of cells from a starting position (x_1, y_1) to subsequent positions (x_i, y_i) can be modelled either (a) on a discrete lattice or (b) in continuous space. Dashed circles show the cell's previous positions with the filled cell representing its final position.

extensions are presented in [117]. It is always useful to consider the differential analogue of agent-based probabilistic models (and vice versa) as different applications may favour different mathematical approaches.

The concept of an isotropic random walk (also known as Brownian motion) provides a flexible basis for a detailed characterisations of cell movements. Remarkably, it can be naturally extended to a wide range of circumstances. For example, cells often have a preferred direction, e.g., in response to differences in their environment. Random walks can be biased by an external source giving preference to movement in a particular direction (a biased random walk or BRW) [109]. This bias can be accounted for by making the step length in the preferred direction longer or by increasing the probability that the preferred direction is chosen. The BRW trajectory shown in Figure 1.7 is biased towards the right (in this case the step length remains the same in every direction but it is more likely that a movement to the right will occur than any other). BRWs can be a suitable model for cells in the presence of biasing chemicals or boundaries [87, 118–120].

As the IRW leads to the diffusion equation, the BRW also has an analogous differential equivalent: the drift-diffusion (or advection-diffusion) equation

$$\frac{\partial n}{\partial t} = D\left(\frac{\partial^2 n}{\partial x^2} + \frac{\partial^2 n}{\partial y^2}\right) - v_x\frac{\partial n}{\partial x} - v_y\frac{\partial n}{\partial y}. \tag{1.2}$$

The first term in the equation represents the diffusive part of the movement as in Eq (1.1) with the second and third terms capturing the drift (bias) in the x and y directions at the speeds v_x and v_y, respectively. Diffusion equations such as Eq (1.1) and Eq (1.2) are not only used in migration models, but also for descriptions of chemotaxis [121] and ecological models [122].

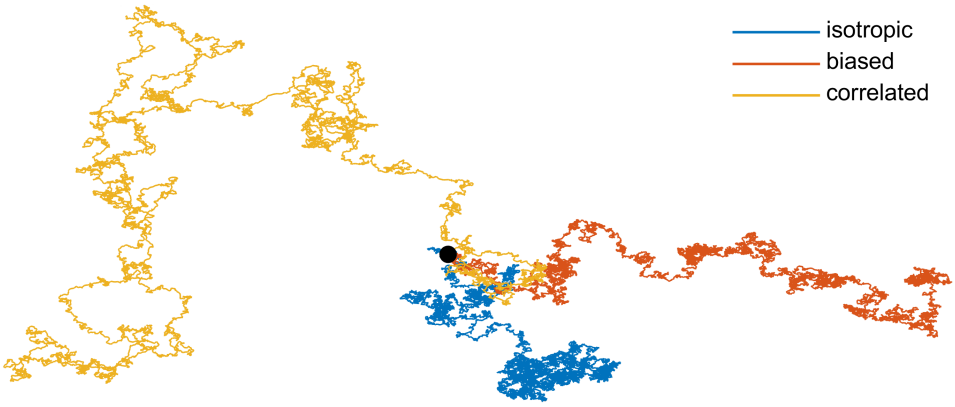

Figure 1.7: Simulated trajectories of isotropic (blue), biased (orange) and correlated (yellow) random walks. All trajectories start from the same position (the black circle). The biased trajectory is extended towards the right and the correlated trajectory has relatively long parts with large directionality. Simulation details: all walks have the same number of steps of the same length. For the isotropic walk the position is updated as $\mathbf{x}_{i+1} = \mathbf{x}_i + \xi_i$, with ξ_i drawn from Brownian noise. The biased random walk includes an additional directional bias to the right with positions updated as $\mathbf{x}_{i+1} = \mathbf{x}_i + \xi_i + 0.25$. The correlated random walk has ξ_i drawn from fractional Brownian noise with a Hurst exponent of 0.65.

Another generalisation of the IRW is a correlated random walk (CRW) where the direction of each step depends on the previous step directions. The walk can be persistent, where the next step is more likely to be close to the direction of the previous step, or anti-persistent, where the next step is more likely to be in the opposite direction. The example CRW trajectory shown in Figure 1.7 is persistent, the cell prefers to keep going in the direction in which it has just been, and so the trajectory ends up more directed than for isotropic movement. CRWs often occur in cell kinematics in the absence of external biases [123–126]. More generally, animal movements often take the form of a CRW (e.g., cabbage butterflies [127], caribou [128] and seals [129]).

Individual cell movements have been explored through direct experiments with hPSCs (in particular hESCs) and analysed within the random walk framework described above [87–89]. The movement of single hESCs can be accurately described as an isotropic random walk when the cells are in isolation, i.e., more than approximately 150 μm away from any neighbouring cells [87]. We apply this analysis framework to our own experimental results for isolated and pairs of hESCs in Chapters 2 and 3.

1.3.2 Differential equations

Differential equations (DEs) are a fundamental part of the mathematical descriptions for all biological systems. They describe the relationship between the properties of the cells, their environment, and their rates of change in space and time. In Section 1.3.1 we encountered two DEs from the theory of random walk models: the diffusion equation and the drift-diffusion equation. This section introduces some further common DEs used in mathematical biology, discusses the differences between deterministic and stochastic DEs, and presents some pertinent examples relevant to stem cell biology, such as the logistic equation.

Fundamental equations

DEs describe rates of change, for example, the rate of temporal change is the time (t) derivative. For example, the rate of change of the number of cells (N) is written as dN/dt (or $\partial N/\partial t$ if N depends on more variables than just t). For a population of N cells, with a fraction r of cells dividing in unit time, the DE describing the evolution of the population is $dN/dt = rN$, meaning that the instantaneous differential (increment) of N per unit time is given by rN. Negative values of r describe a dying population and positive value correspond to its proliferation. This equation has the solution of the form $N = N_0 e^{rt}$ where N_0 is the initial size of the population (at $t = 0$, the starting time of the evolution). The evolution of N over time for different initial population sizes and different growth rates is shown in Figure 1.8. This equation can be further enriched by including various effects that influence the cell division, as we discuss below.

There are two major types of differential equation: deterministic and stochastic. Deterministic DEs have no randomness in their parameters (such as the growth rate r) and so have the same solution for a given initial state. Stochastic DEs (or SDEs) include randomness in the parameters, which results in different solutions even for the same initial state. Real-life biological scenarios often include some element of randomness, and it is for this reason that SDEs are often chosen as better models. We will use the deterministic and stochastic logistic equation, used to model population growth, to illustrate the differences between deterministic DEs and SDEs.

The logistic growth model is fundamental in cell biology and ecology. The equation describes the evolution of population size N with time t as

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right), \tag{1.3}$$

where the growth rate of the population is r and K is the carrying capacity, which represents the maximum population size allowed in a given environment. As N increases,

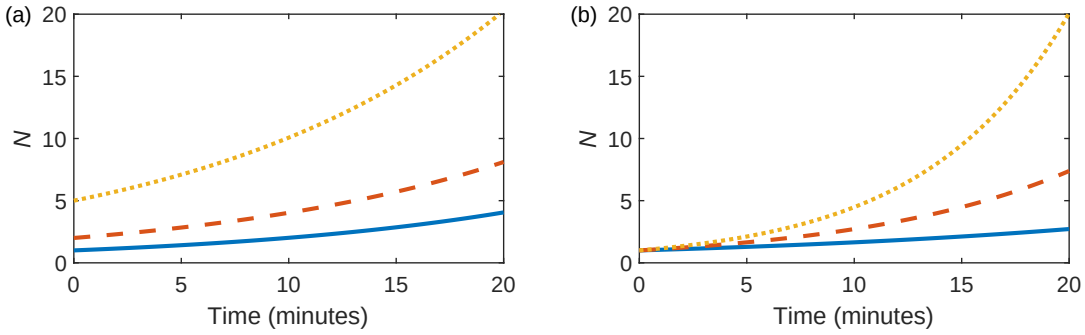

Figure 1.8: (a) The time evolution of a population N following exponential growth, $N_0 e^{rt}$, with $r = 0.06$ cells/minute and $N_0 = 1$ (blue solid), 2 (orange dashed) and 5 (yellow dotted). (b) Exponential growth with $N_0 = 1$ and $r = 0.05$ (blue, solid), 0.1 (orange, dashed) and 0.15 cells/minute (yellow, dotted).

the instantaneous growth rate $r(1 - N/K)$ decreases until the growth stops when $N = K$ and the population settles into a steady state. Examples of a population growing according to Eq (1.3) are shown in Figure 1.10(a). This equation is deterministic, solving the equation again with the same initial condition N_0 gives the same results.

There are other adaptations to the logistic growth model, including time-dependent carrying capacities and time delays where changes in the system parameters affect the population size only after a certain delay. Another common modification used in mathematical biology is the logistic equation with an Allee effect [130–132]. An Allee effect causes reduced population growth rates at low population sizes. The deterministic logistic equation for a population N can be written with an example Allee effect term as

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right)\left(\frac{N - A}{K}\right), \tag{1.4}$$

where A is critical point at which the Allee effect occurs. The effect of the Allee term in Eq (1.4) on both dN/dt and the population N for an example system is illustrated in Figure 1.9. For a weak Allee effect ($A < N_0$), the change in population with time, dN/dt, remains positive for $N < K$ but is significantly suppressed. For a stonger Allee effect ($A > N_0$), dN/dt is negative for $N < K$ and results in the population declining to zero.

The stochastic logistic equation adds randomness to the standard logistic model of Eq (1.3). Some element of randomness, or noise, can be essential in many circumstances. The population then broadly follows the logistic equation, Eq (1.3), but only on average. This approach is based on the concept of averaging over different realisations of the population growth that have slightly different environments and individual cell properties. Each realisation can represent a separate experiment. It is impossible to reproduce the system with perfect accuracy in different experiments. This uncertainty

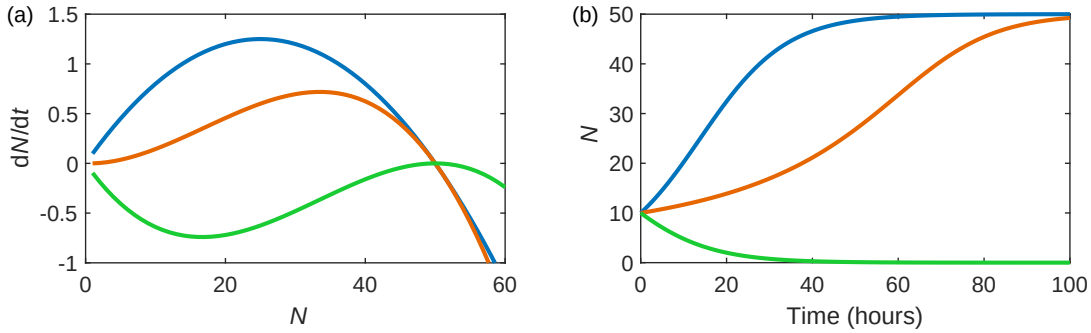

Figure 1.9: The deterministic logistic equation with an initial population size of $N_0 = 10$, $r = 0.1\,/\text{h}$, $K = 50$ and an Allee effect, Eq (1.4), for (a) dN/dt and (b) N with $A = 1$ (orange) and $A = 50$ (green). The deterministic logistic equation with no Allee effect, Eq (1.3), is shown in blue. Each differential equation, Eq (1.4) with varying parameters, was solved numerically using the Euler method.

leads to the randomness. One of the fruitful stochastic logistic models is where the overall rate of change of dN/dt fluctuates:

$$\frac{dN}{dt} = rN\left(1 - \frac{N}{K}\right) + \sigma W_t, \tag{1.5}$$

where the additional noise term is given by σW_t (this is known as the Wiener process, purely random noise analogous to the Brownian motion) with the scaling parameter σ which determines the strength of the noise and W_t which captures the statistical properties of the noise (e.g., Gaussian). Examples of the evolution described by Eq (1.5) are shown in Figure 1.10(b). Although each of the three solutions follow the same trend, i.e., the average behaviour, the noise in the system results in three different evolution trajectories. The scaling σ controls the strength of the random variation; the greater σ the greater the variation in the solutions as shown in Figure 1.10(c).

The randomness of real systems can affect not only the overall rate of growth dN/dt, as in Eq (1.5), but also the intrinsic growth rate r. This source of randomness reflects effects such as variation in the length of the cell cycle from cell to cell. Thus, the intrinsic growth rates now contain a random part, becoming $r + \sigma W_t$. Stochastic logistic equations have been used widely, particularly in ecology and cell biology. In Chapter 6 we consider a novel application of a variation of the stochastic logistic equation, using it to model the levels of pluripotency in a colony.

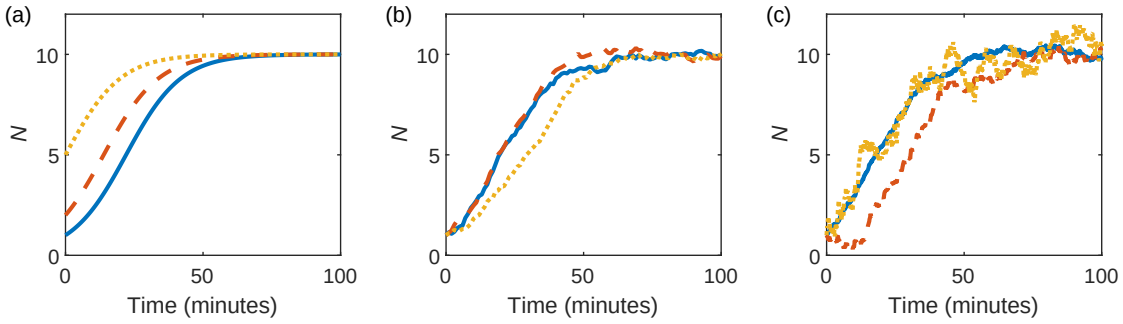

Figure 1.10: (a) The population size growing with time as described by the deterministic logistic equation, Eq (1.3), with the growth rate $r = 0.1$ cells/minute, carrying capacity $K = 10$ cells, and the initial population sizes of one (blue, solid), two (orange, dashed) and five (yellow, dotted) cells. (b) Different realisations of the solutions for the stochastic logistic equation, Eq (1.5), with the noise magnitude $\sigma = 0.1$. The blue solid line in (a) represents the mean value of such realisations. (c) Solutions to the stochastic logistic equation with increasing noise; $\sigma = 0.1$ (blue, solid), 0.2 (orange, dashed) and 0.5 (yellow, dotted). The differential equations were solved numerically using the Euler-Maruyama method (a generalisation of the Euler method for SDEs).

Colony growth models

Population models have been used to understand the process by which blood cells are formed [96], cancer tumours grow [133] and in Chapter 4 of this work, the impact of hPSC colony growth on clonality [82]. Early population dynamics models for stem cells were based on stochastic birth-death processes [91] involving systems of ordinary differential equations [134] similar to the logistic equation discussed above. One of the most popular models for hPSCs allows for dividing and non-dividing cells, and incorporates cell loss through death or differentiation (often referred to as the Deasy model, which is a development of the Sherley model to include cell loss) [135, 136]. The evolving number of cells over time $N(t) = N_{\text{proliferating}} + N_{\text{differentiated}}$ is obtained as

$$N_{\text{proliferating}} + N_{\text{differentiated}} = N_0 \left[\frac{1}{2} + \frac{1 - (2\alpha)^{t/T+1}}{2(1-2\alpha)} - M \right], \tag{1.6}$$

where N_0 is the initial number of cells, α is the mitotic fraction (the fraction of the cells that can divide), T is the cell cycle duration, and M is the number of cells lost to death or differentiation. Since the model accounts for both differentiating and proliferating cells, it can be used to assess the heterogeneity of a cell population and the evolution of mixed sub-populations [135]. The model has been compared with experimental results for muscle-derived stem cells [135] and hESCs [137].

More recently, a new family of growth models has been developed, known as hyperbolastic growth models. These models are designed to capture self-limiting growth

behaviours [138] and have been introduced for both adult and embryonic stem cells [137]. The hyperbolastic growth models provide more flexibility in the growth rate as the population reaches its carrying capacity and have been demonstrated to capture experimental data well [137, 138]. The population N in this case is governed by the differential equation

$$\frac{dN}{dt} = (K - N)\left[r\gamma t^{\gamma-1} + \frac{\theta}{\sqrt{1+\theta^2 t^2}}\right], \tag{1.7}$$

describing the rate of change in the number of cells, dN/dt, starting from the initial condition $N = N_0$ at $t = 0$. The parameters represent the salient properties of the system: K is the carrying capacity of the environment, r is the intrinsic growth rate, the factor $t^{\gamma-1}$ with a certain constant γ is responsible for a decrease ($\gamma < 1$) or increase ($\gamma > 1$) in the cell proliferation rate with time, and allows for further variation in the instantaneous growth rate. This model can be used to describe both proliferation and cell death rates more accurately than Eq (1.6) [137] and helps identify when the growth of cells becomes self-limiting, a biological problem currently not fully understood.

In Chapter 4 we use experimental data to elucidate a population model for hESC colonies to investigate the critical time at which clonality is lost due to physical colony merges [82].

Pluripotency models

Mathematical models of pluripotency are deepening our understanding of how pluripotency is regulated, leading to the optimisation and control of cell pluripotency in the laboratory. Modelling pluripotency and cell fate decisions remains challenging as even clonal cells under the same conditions make different fate decisions, and it remains unclear how much fate choice is led by inherited factors versus environmental factors and intra-cellular signalling [139]. There are several thorough reviews of the computational models of cell fate decisions [140–142].

Here we focus on a few current models that aim to capture the fluctuating pluripotency in individual cells over time. At the single-cell level, the expression of pluripotency transcription factors (PTFs) is highly variable and subject to intrinsic noise [60, 143, 144] leading to colonies exhibiting heterogeneous sub-populations of cells with differing levels of PTF expression [39, 42]. This suggests that the evolution of cell colonies involves a competition between disruptive single-cell influences and regulatory community effects. There is recent evidence that differentiation potential is linked to the length of the G1 phase of the cell cycle [145]. Such heterogeneity is undesirable as it can bias evolution trajectories and lead to spatially disordered differentiation [61–63]. Here we will consider intra-cellular models of pluripotency that focus on the inter-connected dynamics of OCT4, NANOG and SOX2, the main PTFs. A network diagram of the

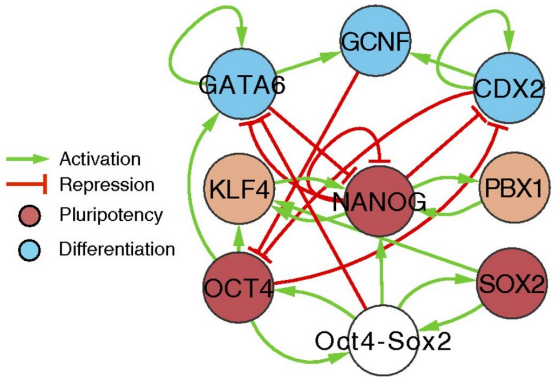

Figure 1.11: A portion of the pluripotency gene regulation network. Each node represents either a transcription factor (with the exception of the protein complex OCT4-SOX2) involved in early stem cell differentiation. Arrows that end at transcription factors represent transcriptional regulations. Activation is shown in green and repression in red. Gene markers found in cells in pluripotent states are coloured in red and orange, and those found for differentiated state phenotypes are coloured in blue. Figure and caption adapted from [146].

regulatory system is shown in Figure 1.11.

A widely used model of the intra-cellular PTF network regulation involves the Hill equation [147, 148] of the theory of chemical reactions. In the simplest form, the concentration of a reacting substance X activated by the transcription factor Y is described as

$$\frac{dX}{dt} = r\frac{Y^n}{K^n + Y^n}, \tag{1.8}$$

where r is the maximum transcription rate, K is the value of Y at which the rate of change of X is reduced to $r/2$ (the so-called equilibrium constant) and n (the Hill coefficient) characterises the sensitivity of the dynamics of X to the magnitude of Y: the larger is n, the higher is the sensitivity. This equation can be generalised and extended to describe a complex network of mutually regulating chemical reactions. The model of [147] describes the rate of change of each transcription factor in the network with the equation for the concentration of OCT4 (O) in relation to NANOG (N), SOX2 (S) and OCT4-SOX2 (θ) given by

$$\frac{dO}{dt} = \frac{\eta_1 + a_1 A_+ + a_2\theta S + a_3\theta N}{1 + \eta_2 + b_1 A_+ + b_2\theta + b_3\theta N} - \gamma O - k_1 OS + k_2\theta, \tag{1.9}$$

where η_1 and η_2 are the basal transcription rates, A_+ is a signal positively regulating the levels of OCT4, γ represents the decay rate, k_1 and k_2 are the kinetic constants and $a_{1,2,3}$ and $b_{1,2,3}$ are the constants related to the free binding energies. The large number of parameters that appear in the resulting system of four equations, which are difficult to estimate accurately, diminishes the advantages of this approach [140, 149].

The NANOG regulation of OCT4 and SOX2 has also been considered using similar differential equations [150]. This model considers two different regimes for NANOG expression: stochastic fluctuations and deterministic oscillations. For stochastic fluctuations alone, a Gaussian white noise is included in the differential equation for NANOG concentration (N) in relation to OCT4-SOX2 (θ)

$$\frac{dN}{dt} = \frac{s_4 N^{\gamma_N}}{k_4 + N^{\gamma_N}} + \frac{s_3 \theta^{\gamma_O}}{k_3 + \theta^{\gamma_O}} - d_N N + N\zeta(0, \sigma), \tag{1.10}$$

where the coefficients s_3 and s_4 denote the transcription rates from the OCT4-SOX2 complex and self-influence, respectively, K_3 and K_4 are the equilibrium constants, γ_N and γ_O are the Hill coefficients and ζ denotes the random contribution proportional to the concentration of NANOG N whose magnitude is controlled by the standard deviation of the Gaussian noise σ. The term with d_N describes the natural degradation of NANOG.

For certain parameter values, the NANOG concentration tends to stabilise at either a low or high level but the random perturbations induce transitions between the two states. In the other regime, which reproduces oscillatory behaviours, the governing equation includes a competitive repression of NANOG due to a transcriptional repressor, denoted X,

$$\frac{dN}{dt} = \frac{s_4 N^{\gamma_N}}{k_4 + N^{\gamma_N} + s_6 X^{\gamma_X}} + \frac{s_3 \theta^{\gamma_O}}{k_3 + \theta^{\gamma_O}} - d_N N + N\zeta(0, \sigma), \tag{1.11}$$

while the repressor dynamics are governed by a similar equation,

$$\frac{dX}{dt} = \frac{s_5 N^{\gamma_N}}{k_5 + N^{\gamma_N}} - d_X X. \tag{1.12}$$

The oscillatory system is characterised by a limit cycle where the system converges to a predictable, periodic state, independent of initial state. Although indistinguishable on the cell population level, the two mathematical models suggest quite different inherent NANOG variation, and further experimental testing using intra-cellular temporal NANOG data is required to assess the applicability of these models. There are many other approaches to the PTF modelling using differential equations [151–156].

Models of this kind can reproduce a wide range of complex behaviours and potentially include a large number of various factors that affect the bio-chemical networks. However, the usefulness of this approach is strongly limited by the fact that the numerous parameters that appear in the resulting equations are difficult or impossible to determine with any degree of accuracy and generality. On the other hand, the very fact that the number of factors that affect the PTF dynamics is so large, opens up the possibility to describe the bio-chemical regulation as a random system where the complex network of interacting positive and negative feedback loops leads to dynamics

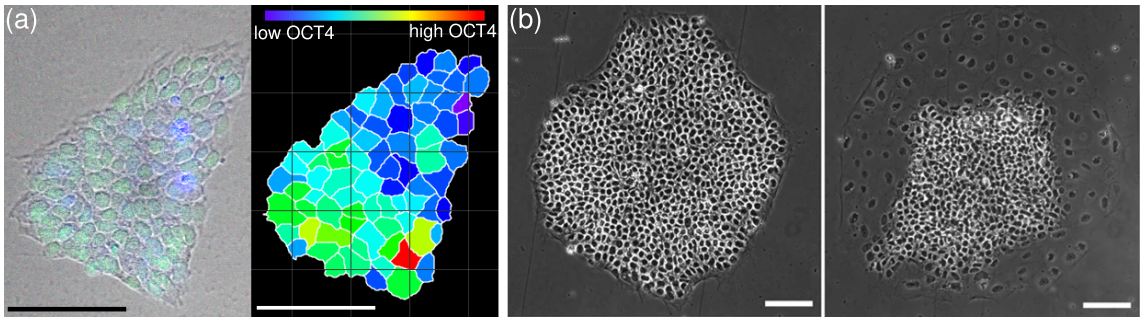

Figure 1.12: (a) A microscopy image of a hESC colony at 72 h after seeding, alongside a colour-coded version of the same colony quantifying the level of expression of OCT4. Red represents the highest pluripotency with purple representing the lowest. (b) Phase images of hESC colonies before and after BMP4 addition showing a band of differentiation. Figure adapted from [61]. The scale bar represents 100 μm in all panels.

captured by a few universal parameters. The situation is not dissimilar to the description of a complex random motion in terms just a single parameter, the diffusivity D as described in Section 1.3.1. In other words, beyond a certain level of complexity, the dynamics of bio-chemical networks is amenable to the stochastic chemistry approaches.

Pluripotency also shows characteristic spatial variation on the colony scale. Preliminary experiments monitoring the levels of OCT4 in single hESC founder cell colonies at 72 h post seeding show that pluripotency is clustered, with highly pluripotent cells grouped together, shown in Figure 1.12(a).

The differentiation of hPSCs also shows distinctive and apparently systematic spatial patterning [61–63]. The spatial distributions of the pluripotency marker SOX2 and the differentiation marker AP2α suggest that the differentiation occurs preferentially at the colony periphery in a band of a constant width, about 150 μm independent of colony size as shown in Figure 1.12(b) [61] and illustrated schematically in the rightmost panel of Figure 1.2. The differentiated cells originate in the outer third of the colony and remain near its boundary as the colony grows. This provides important information for modelling the spatial patterning of the pluripotent state.

The spatial segregation of differentiated cells can be partially captured in a model of cell movements in a colony which is based on the balance of the intra-cellular and extra-cellular mechanical forces [157]. This is known as a mechanical bidomain model and was first developed to describe the elastic behaviour of cardiac tissue [158]. In agreement with the experimental results [61–63], the model predicts that differentiation and traction forces occur within a few length constants of the colonies edge. The length scale is determined by the intra-cellular and extra-cellular cell rigidities (shear moduli) and the coupling of the internal and external mechanical forces. The model assumes that differences in displacement between the pluripotent and differentiated cells are

responsible for the mechano-transduction (chemical processes through which cells sense and respond to mechanical stimuli). Equations for the intra-cellular (u_r) and extra-cellular (w_r) displacements, functions of the distance from the colony centre (r), have the form

$$4\nu\left(\frac{\partial^2 u_r}{\partial r^2}+\frac{1}{r}\frac{\partial u_r}{\partial r}-\frac{u_r}{r^2}\right)=K(u_r-w_r), \tag{1.13}$$

$$4\mu\left(\frac{\partial^2 w_r}{\partial r^2}+\frac{1}{r}\frac{\partial w_r}{\partial r}-\frac{w_r}{r^2}\right)=-K(u_r-w_r), \tag{1.14}$$

where ν and μ are the intra-cellular and extra-cellular shear moduli, respectively, and K is their coupling constant. The difference between the intra and extra-cellular displacements follows as

$$u_r-w_r=-\frac{T\sigma}{4\nu}\mathrm{e}^{\frac{r-R}{\sigma}}, \tag{1.15}$$

where T is a uniform stress caused by the growth and crowding of cells, $\sigma=\sqrt{4\nu\mu/[K(\nu+\mu)]}$ is the characteristic length and R is the colony radius. As shown by this relation, the difference is very sensitive to the distance $r-R$ to the colony boundary and rapidly decreases away from the boundary at the length scale σ which is independent of the colony radius R.

This model assumes that the difference between the intra-cellular and extra-cellular forces is the primary driver (or evidence of) the differentiation. This model could be further developed to allow for an irregular colony shape. It is also worth investigating whether the cell growth, represented by the tension T, is a function of u_r-w_r alone (as assumed in the model), since hESC observations suggest distinct actin organisation and greater myosin activity (both proteins that facilitate cell movement) near the colony edge, implying that T could be non-uniform [61].

Further intra-cellular PTF monitoring experiments using fluorescent staining (such as [41]) are required to explore the effects of different geometries and growth plate boundaries, varying cell lines and colony sizes on pluripotency. In addition to those already mentioned, thorough reviews of other mathematical models of pluripotency are available [92, 159], along with a review of computational modelling of the fate control of mouse embryonic stem cells, with many models transferable to hPSCs [140]. The mathematical exploration of pluripotency promises a better understanding of pluripotency regulation and differentiation within colonies.

In this thesis we explore the regulation of the pluripotency transcription factor OCT4 in Chapters 5 and 6. We first rigorously quantify the nature of the intra-cellular regulation of OCT4, and then explore techniques for mathematical modelling, including a variation of the stochastic logistic equation discussed previously.

1.3.3 Agent-based modelling

The understanding of larger groups and colonies of hPSCs is complicated by the fact that they involve both collective and individual behavioural effects [71, 160]. For example, the coordinated migration of large numbers of hPSCs *in-vivo* is essential in tissue generation [69] and wound healing [70], and as discussed in Section 1.3.2, pluripotency shows spatial patterning on the colony scale [61–63]. A flexible approach to modelling large colonies of cells with allowance for their complicated multi-scale behaviours is the agent-based modelling technique.

Agent-based models (ABMs, sometimes referred to as individual-based models or IBMs) consist of individual components, called 'agents', which are characterised by their locations and behave according to a set of rules that reflect their response to the environment and internal processes. This type of model is based not on differential equations where the equation parameters reflect our knowledge of the cell properties, but rather on a set of explicit rules for the evolution of each agent in response to various stimuli. It is often convenient to think of an agent as a single cell although it can represent a group of cells if appropriate. The governing rules are often assumed to be probabilistic. The rules of behaviour usually include the probability of movement in any of the allowed directions (the probabilities often depend on the state of the neighbouring cells) and changes in the agent state depending on time or the neighbouring agents (implementing, for example, the cell cycle). ABMs are commonly used in the studies of collective behaviours in mathematical biology and population dynamics not only for cells but also for schools of fish [161] and crowds of people [100, 162].

Each ABM agent can represent an individual cell in a colony. The set of rules that describe the behaviour of the cells and the colony as a whole can include a wide variety of external and internal factors leading to complex and sophisticated models. For example, each cell's individual migration could occur as an isotropic random walk (in fact, the random walk models described in Section 1.3.1 can be implemented as ABMs). Further rules can then be added to describe the cell cycle and the cell interactions with its neighbours. In their simpler incarnations, ABMs contain explicit rules deciding each agent's evolution, whereas further sophistication is achieved by deriving parameters that enter those rules from the solution of a system of differential equations. The idea of agent-based modelling is to learn something about the whole system (i.e., macroscale and collective behaviours) from various assumptions about the individual (or microscale) behaviour.

One of the key advantages of ABMs is their vast flexibility; it is simple to add more agents, change the governing rules and add or remove layers of complexity in the cell behaviour. They have strong predictive power having been used to predict the emergent behaviour of cell interactions [163], model self-organisation within colonies [164] and

describe stem cell driven tumour growth [165]. A disadvantage of ABMs is their computational cost which increases rapidly with the number of the agents and the sophistication of rules that govern their behaviour. It is also difficult to develop simplified analytical models to interpret and generalise the results of ABM simulations.

ABMs for cell migration

ABMs have been developed to incorporate the coordinated migration and collective behaviours of stem cells within colonies, but the challenges remain to fully capture the experimental behaviours, especially collective aspects and cell migration in three dimensions. These agent-based migration models are often combined with models of colony growth and proliferation [84, 166].

hPSCs show coordinated intra-colony movements which cease upon differentiation [167]. Cell movement speed varies within colonies, with higher average speed at the periphery and lower in the central region [57]. Recently, a two-dimensional (monolayer) individual-based stochastic model was developed of cell migration, cell-cell connections and cell-substrate connections which captures well these experimental observations [84]. The model is formulated in terms of the energies associated with the cell-cell and cell-substrate connections. Any energy released by breaking and forming such connections drives cell migration in one of the eight directions on a square lattice along the lattice sides and diagonals. The direction of movement is determined at random according to a probability related to the cell's energy and a spatial weighting which favours the sideways rather than diagonal displacements (as described in [168]). Cell proliferation and quiescence are also included. This model suggests that cell division is a leading factor in the increased mobility at the colony edges. This is a promising approach for the studies of complex behaviours of hPSCs and planning of bio-processing experiments.

Models of cell cultures in three dimensions are developing actively as they are required to achieve a more realistic understanding of *in-vivo* behaviours of hPSCs [169] and *in-vitro* engineering of tissues on three-dimensional scaffolds [170]. A review of current mathematical models for organoid cultures is given in [171]. Other recently developed models also provide a good starting point for such simulations, such as the open-source PhysiCell model originally devised for cancer cells but transferable to other cell types including hPSCs [108]. The model implements cell movement by defining a persistence time, a migration speed and a migration bias, allowing for a range of cell motions from purely random to deterministic. The model also includes mechanical interactions between neighbouring cells. An example of the outcome of such simulations for a hanging drop spheroid (a type of 3D cell culture) with deterministic and stochastic necrosis is shown in Figure 1.13.

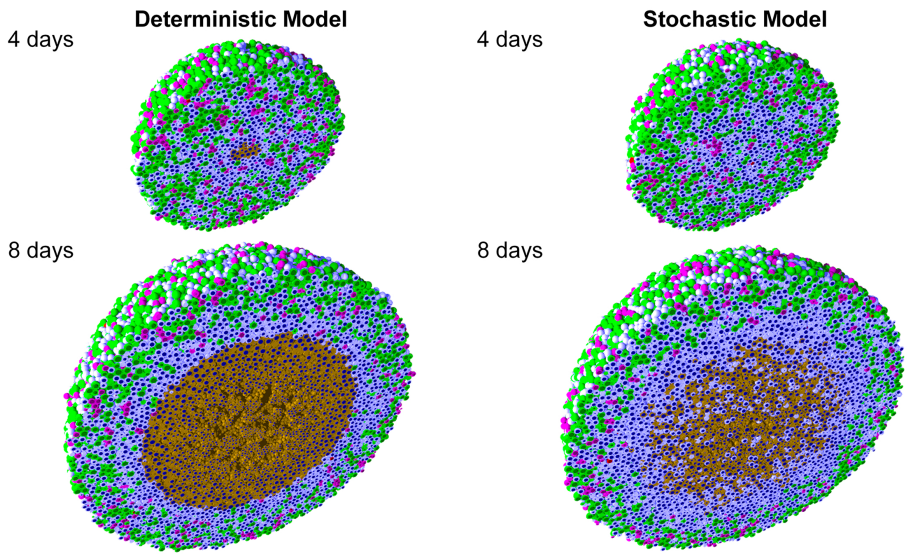

Figure 1.13: Hanging drop spheroid simulations with deterministic necrosis and stochastic necrosis. Different cell states are colour coded. These simulations predict the emergence of a necrotic core microstructure arising from cell-scale mechanical interactions (adapted from [108]).

Modelling cell movement on a three-dimensional discrete lattice is widely used, e.g., for modelling the mesenchymal stem cell tissue differentiation [172] and tumour growth driven by cancer stem cells [165]. Some models allow many lattice nodes per cell representing a cell as a spatially extended object (as in cellular Potts models) [173–175]. There is also a range of agent-based continuous models where cell movement is not restricted to a grid but occurs continuously in any direction as illustrated in two dimensions in Figure 1.6 [176, 177]. Here the movement is described in terms of forces or potentials, with positions obtained from differential equations of motion for each cell. In centre-based models, each cell is represented by a simple geometrical object, such as a circle, whereas vertex models describe a cell via connected nodes [178].

The flexible nature of ABM can allow the cells' changing morphology to be included. For example, a model for mesenchymal stem cells [166] includes the random formation, elongation and retraction of pseudopodia resulting in forces that lead to cell movement. This model can reproduce the spatio-temporal organisation of cells at a quantitative level and emphasises the importance of cell-cell interactions in tissue formation. In its published form, this model shows stronger inertial and accelerated movements than experimental results [111]. How much of this discrepancy is due to differences in cell type and culturing conditions needs to be investigated further to clarify the model's applicability to different experiments.

Colony proliferation

The simplest ABMs incorporate cell proliferation probabilistically, with the division time for each cell drawn at random from a suitable probability distribution [84]. Others go a step further by moving cells through each cell cycle phase according to timings based on experimental data [163] or as cell volume increases [166]. Sometimes divisions do not occur; this probabilistic nature of self-renewal can be incorporated when the end of the cell cycle is reached [179].

The ABMs discussed in the previous section are implemented on a fixed lattice, so that the agent positions change in jumps from one lattice node to another. An alternative is to allow for an irregular shape and variable size of a region allocated to each agent (cell). In one implementation of this approach, each cell is represented as a vertex in a graph whose edges are straight lines joining each cell with its nearest neighbours, as shown with dashed lines in Figure 1.14(a). To define the neighbourhood of each cell, a perpendicular line is drawn through the middle of each edge to form a closed polygon with a cell inside. Such a division of the plane into polygonal cells (Voronoi cells) is called the Voronoi tessellation. A Voronoi cell consists of all points that are closer to its agent than to any other. Similar procedure in the three-dimensional space results in the space partition into convex polyhedrons. The Voronoi cells are not uniform in shape and have a variable number of sides depending on the number density of the cells. The tessellation can also be constructed to represent experimental images using the geometrical cell centroids or their nucleus positions [104].

Voronoi tessellation has been used to model monolayer adult stem cells in intestinal crypts [180, 181] and is now being transferred to hPSCs. The model uses an agent-based approximation in which each cell is represented as a Voronoi tessellation of the space [180, 182]. The domain grows according to the inter-cellular pressure that builds up due to mitotic divisions in the colony. The cell-cell interactions are described by a force derived from an elastic potential given for a cell labelled with a number i by

$$V(\mathbf{r}_i, t) = \frac{k_v}{2}\left[\alpha_i(t) - \overline{\alpha}_0(t)\right]^2 + \frac{k_c}{2}\left[\mathbf{r}_i(t) - \mathbf{r}_{0i}(t)\right]^2, \tag{1.16}$$

where k_v and k_c are the elastic constants, α_i the area of the i'th cell, $\overline{\alpha_0}$ is the equilibrium cell area and r_i is the initial positions of the i'th cell which does not necessarily coincide with its centroids position r_{0i}. The first term in the right-hand side of Eq (1.16) reflects the tendency for the cells to have similar sizes, whereas the second term controls the cell shape. The gradient of the potential combined with a drag force gives the total force acting on each cell. A simulated colony undergoing a cell division is shown in Figure 1.14(b). Cells in the middle of the colony experience a higher pressure and show mitotic arrest, i.e., they do not divide.

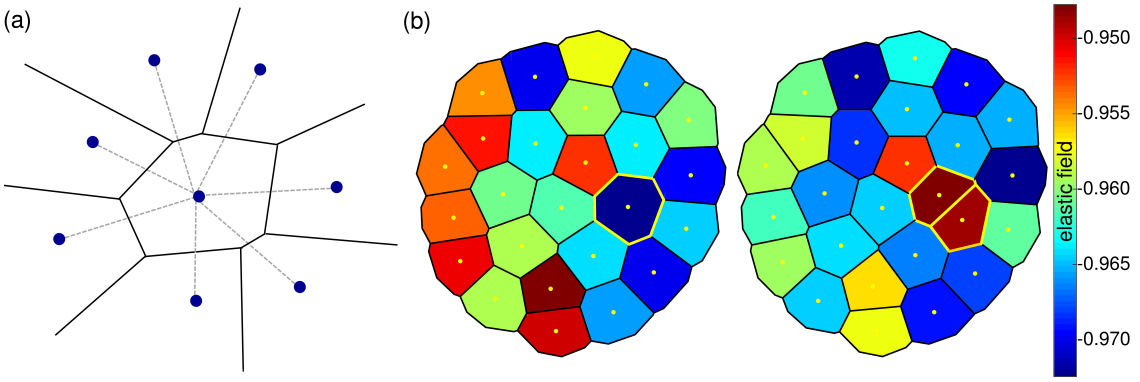

Figure 1.14: (a) The Voronoi diagram (with edges shown black solid) illustrating how colony area is divided into tessellated cells around the agents (blue circles) with the nearest neighbours joined by dashed lines. (b) Voronoi tessellation to simulate a proliferating hPSC colony. The cells divide and give rise to two daughter cells, see highlighted cells outlined in yellow. The colour bar shows the elastic field in Eq (1.16). Model and colony images by Dr Sirio Orozco-Fuentes.

Spatially modelling each individual cell in a colony in this way raises an important question about the physical process involved in cell division: how does the colony rearrange to make space for new cells? In Voronoi tessellation models [180, 182] the cell rearrangement is controlled by the potential forces from the neighbouring cells and the crypt wall. In most square- or hexagonal-lattice models, one daughter cell is placed in the same position as the mother cell while the other is put into a neighbouring position, chosen at random [183], isotropic mitosis. If there is no free position available next to the dividing cell, the neighbouring cells are re-arranged into the free spaces available stochastically until there is a free space next to the dividing cell [84] or, if this is impossible, mitosis is suppressed (quiescence) [172, 184]. In cellular Potts models, division occurs when cells reach a target volume, with surrounding cells being pushed outwards [173].

The cell proliferation in a colony is sensitive to various environmental factors. There is evidence that dense cell packing reduces cell proliferation [58], which has been captured in a model showing preferential cell division at the colony edge [84]. Self-organisation of cells has also been observed, where the newly divided (smallest) cells cluster together in patches, separated from larger cells at the final stages of the cell cycle [104]. This segregation by cell size allows neighbouring cells to move and swap places as the colony grows and could directly influence cell-to-cell interactions and community effects [104].

The PhysiCell model [108] described above allocates each cell a volume which varies with the cell cycle. The daughter cells have half the volume of their parent and are placed next to their parent cell position. A combination of this model and hPSC-specific spatial models is a promising approach to modelling the structure of three-dimensional

colonies. Spatial models of hPSCs become increasingly complex with colony size, and it is difficult to incorporate collective migratory effects. These models have already advanced our understanding of the growth of cancer tumours [185] and wound healing [186].

1.3.4 Discussion

Mathematical models of cell behaviour are essential in developing non-invasive predictive tools in stem cell biology. Their purposes include the identification of inherent and universal systematic behaviours, determination of the key factors that control the cell behaviour, and quantitative comparison of experimental results obtained under different biological conditions.

The complexity of hPSC systems that involve a plethora of chemical, mechanical and biological effects that act at a range of spatial and temporal scales aggravates the modelling and understanding of the intra-cellular, inter-cellular and colony-scale behaviours. A systematic approach to the modelling is essential, and models isolating a limited range of the key properties have often been more successful and useful in interpreting experimental results [90]. For example, focussed migration models (Section 1.3.1) have led to a greater understanding of the behaviour of isolated cells [87–89] and the movement of cells within colonies [84, 166]. The population models for colony proliferation (Section 1.3.2) have been advanced, e.g., to include cell divisions and apoptosis, providing a distinct computational advantage over more complex spatio-temporal models and allowing investigations into the impact of colony expansion on clonality [82]. Spatio-temporal ABM models require more computational power, but have been used to capture cell regeneration within intestinal crypts [180, 181] and tumour growth [185] (Section 1.3.3).

Many current efforts focus on modelling cell pluripotency and cell fate, as applications of hPSCs require greater control over pluripotency and differentiation trajectories. The stochastic nature of pluripotency at the single-cell level [60, 143, 144], along with recent studies of the fluctuations of PTFs throughout colonies [41] and spatial patterning of differentiation [61–63] are being used to inform the development of models of pluripotency and cell fate.

The comparison of models with experiments represents a separate and important aspect of the modelling [97]. Model refinement should be based on a two-way interaction with experiments; model parameters need to be better informed by experimental results. Mathematical models should be used to influence experimental design to achieve a deeper interpretation of the experiments, extending beyond relatively simple descriptive presentations. Models of this kind have already helped provide an insight into tissue formation, wound healing, tumour growth and the reprogramming of hiPSCs and will no doubt continue to do so as these models progress.

1.4 Thesis overview

In this thesis we use the essential combination of experimentation, quantitative analysis, and mathematical modelling to elucidate characterisations of some of the key biological properties of hPSCs. We emphasise that this interdisciplinary approach is an exciting and rewarding method for deepening our understanding of stem cell systems and encourage a continued discussion between biologists and mathematicians working towards a common goal.

In the subsequent chapters, we use statistical analysis of experimental data to inform the development of mathematical models for the key behaviours of hPSCs: cell kinematics, colony growth and pluripotency.

- Chapter 2: Behaviour of isolated hPSCs
  We use experimental data to inform the description of the kinematics of single hESCs under a random walk framework and consider the effects of CellTrace™ on motility.
- Chapter 3: Behaviour of pairs of hPSCs
  We extend the experimental analysis from Chapter 2 to describe the kinematics of pairs of hESCs.
- Chapter 4: Growing colonies of hPSCs
  We develop a stochastic model for the proliferation of hPSC colonies using experimental data and explore the impact of colony growth on clonality.
- Chapter 5: Characterising pluripotency
  Using available experimental data, we quantify the intra-cellular regulation of the pluripotency transcription factor OCT4.
- Chapter 6: Modelling pluripotency
  We explore a range of stochastic techniques for mathematically modelling the temporal regulation of OCT4.
- Chapter 7: Future steps
  We review the limitations of our results and explore the next steps in models of hPSCs.
- Chapter 8: Results summary
  We concisely present the main findings from this work.

Contributions and acknowledgements

The ambitious inter-disciplinary nature of this work was possible thanks to numerous experimental experts. Experiments 1, 2 and 4 discussed in Chapters 2, 3, 7 were carried out at the Institute of Biosciences at Newcastle University by Dr Irina Neganova, with assistance from Dr Valeria Chichagova, Dr Sanja Bojic and Dr Sirio Orozco-Fuentes. Microscopy imaging took place at the Bioimaging Unit at Newcastle University, conducted by Dr Alex Laude. Data collection was conducted by the author, with image analysis training on specialised software provided by Dr Alex Laude. Professor Majlinda Lako, Dr Irina Neganova and Dr Sirio Orozco-Fuentes provided invaluable biological interpretations of the mathematical results. Dr Sirio Orozco-Fuentes developed the agent-based model, shown in Figure 1.14.

Publication list

The following publications are presented in detail within this thesis:

- [88] L E Wadkin, L F Elliot, I Neganova, N G Parker, V Chichagova, G Swan, A Laude, M Lako, and A Shukurov. Dynamics of single human embryonic stem cells and their pairs: a quantitative analysis. *Scientific Reports* 7, 570 (2017).
- [89] L E Wadkin, S Orozco-Fuentes, I Neganova, G Swan, A Laude, M Lako, A Shukurov, and N G Parker. Correlated random walks of human embryonic stem cells in-vitro. *Physical Biology* 15.5 (2018).
- [82] L E Wadkin, S Orozco-Fuentes, I Neganova, S Bojic, A Laude, M Lako, N G Parker, and A Shukurov. Seeding hESCs to achieve optimal colony clonality. *Scientific Reports* 9, 15299 (2019).
- [85] L E Wadkin, S Orozco-Fuentes, I Neganova, M Lako, A Shukurov and N G Parker. The recent advances in the mathematical modelling of human pluripotent stem cells. *Springer Nature Applied Sciences* 2, 276 (2020).
- [72] L E Wadkin, S Orozco-Fuentes, I Neganova, M Lako, A Shukurov and N G Parker. An introduction to the mathematical modelling of iPSCs. In preparation. Induced Pluripotent Stem Cells: Novel Concepts, Elsevier (2020).
- [187] L E Wadkin, S Orozco-Fuentes, I Neganova, M Lako, R A Barrio, A W Baggaley, A Shukurov and N G Parker. OCT4 expression in human embryonic stem cells: spatio-temporal dynamics and fate transitions. arXiv e-prints, page arXiv:2004.09133 (2020).

The following paper, although not discussed in detail, is directly related to and complements the contents of this thesis:

- [104] S Orozco-Fuentes, I Neganova, L E Wadkin, A W Baggaley, R A Barrio, M Lako, A Shukurov, and N G Parker. Quantification of the morphological characteristics of hESC colonies. *Scientific Reports* 9, 17569 (2019).

Chapter 2

Behaviour of isolated hPSCs

Migration is a key behavioural property of hPSCs which affects cell-cell interactions, the pluripotent state, and colony clonality [21, 57, 67, 82, 87, 188]. We begin this chapter by introducing the mechanics of cell migration and then use experimental data to characterise the behaviour of isolated hESCs within the random walk framework described in Section 1.3.1.

We quantify the movement of single cells as a first step to understanding how a hPSC colony is formed and provide a detailed statistical analysis of their kinematic behaviours, including their speed, survival, division time, directionality, distance travelled, and diffusivity. It is a common experimental practice to treat cells with CellTrace™ to allow the tracking of cell generations. We show that this negatively affects the cell motility. These results are also presented in [88] and [89].

2.1 Mechanics of cell migration

There are many different types of active and spontaneous cell motion, e.g., swimming, gliding, crawling and swarming, detected in both prokaryotic and eukaryotic cells [189, 190]. The favour of one mechanism over another depends on the environment and the balance of achieved displacement versus energy expenditure. Cell motility and migration are essential in many biological processes including the development, morphogenesis and regeneration of multicellular organisms, wound healing, tissue repair and angiogenesis [68, 191–195]. Anomalous cell migration can cause developmental abnormalities, tumour growth, neuronal migration disorders and the progression of metastatic cancer [196–198].

Motility is an intrinsic property of hPSCs and they can increase their migratory activity under certain conditions [199]. Understanding the form of cell trajectories provides important insights into diverse cell motility modes and helps to design and interpret experiments. For example, understanding the role of cell migration in metastatic cancer has led to new treatments which modify signalling pathways and alter cell morphology to reduce cell motility [200, 201].

Adaptations in cell morphology facilitate migration. hPSCs and many other eukaryotic cells in culture are anchorage dependent, i.e., they adhere to the substrate surface and spread their cytoskeleton to attach to the surface. They then achieve motion through the coordinated and cyclic reorganisation of the actin cytoskeleton, which determines their speed, direction, and trajectory [53, 202]. The main structures that define the leading edge on a migrating hPSC are referred to as pseudopodia; these protrusions also sense

external cues. A single cell migrating using pseudopodia formation is shown in Figure 2.1. Several types of protrusive pseudopodia structures have been characterised, which mainly differ in the organisation of actin [203]. Analysis of the formation of pseudopods has shown that cells extending pseudopodia which then split into two to allow a change of direction exhibit strong persistence and small turning angles [126, 204].

It is suggested hPSCs perform an unbiased random walk when they are farther than 150 μm apart, with cells closer to one another exhibiting coordinated motion [87]. The cells that are closer to each other move in a more systematic, directed manner, and display a higher ability to form colonies arising from more than one founder cell, suggesting that the separation distance of hPSCs at the start of colony formation and their migration parameters are important for clonal expansion and thus have an impact on pluripotent phenotype and status of a colony as a whole [21, 87].

Unregulated cell migration *in-vitro* can cause clonality loss as the cell population grows, undesirable when a genetically identical clonal population is required [21, 54, 55, 87]. Furthermore, anomalous cell migration has been linked to deviations in the undifferentiated state of hiPSCs [57]. A thorough understanding of the migration of hPSCs is needed to optimise *in-vitro* clonality and facilitate the development of therapies for migration related disorders.

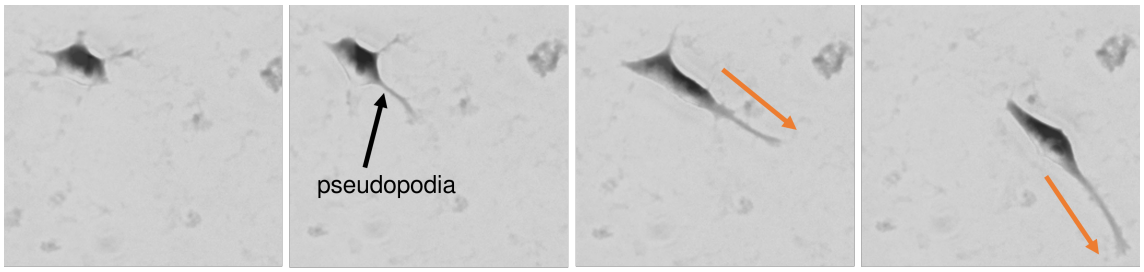

Figure 2.1: A single cell forming a pseudopodia to lead migration. The movement of the cell is in the direction of the leading edge (shown by the orange arrows).

2.2 Individual hPSCs

Unconstrained cell migration on a two-dimensional plane *in-vitro* can often be described as a two-dimensional random walk [109]. We set out the mathematical framework of random walks in Section 1.3.1. In this section we quantify the characteristic kinematics of individual hPSCs and assess the applicability of isotropic and correlated random walk models to their migration using time-lapse imaging experiments. Through this methodology we explore the response of the cell dynamics to CellTrace™ treatment and identify systematic features of the cell behaviour.

Experiment 1 Carried out at the Institute of Biosciences at Newcastle University by Dr Irina Neganova. Data collection and image analysis by the author.

Single human embryonic stem cells were plated at low density, $n_0 = 1500\,\text{cells/cm}^2$. Time-lapse images were taken every 15 minutes and the cell positions and cell fate (death or division) were tracked. A second set of single cells were tracked under the same conditions with the addition of a CellTrace™ Violet Dye, a common biological marker for tracking cell generations. Example microscopy images are shown in Figure 2.2. Further details of the experimental process are given in Appendix A.

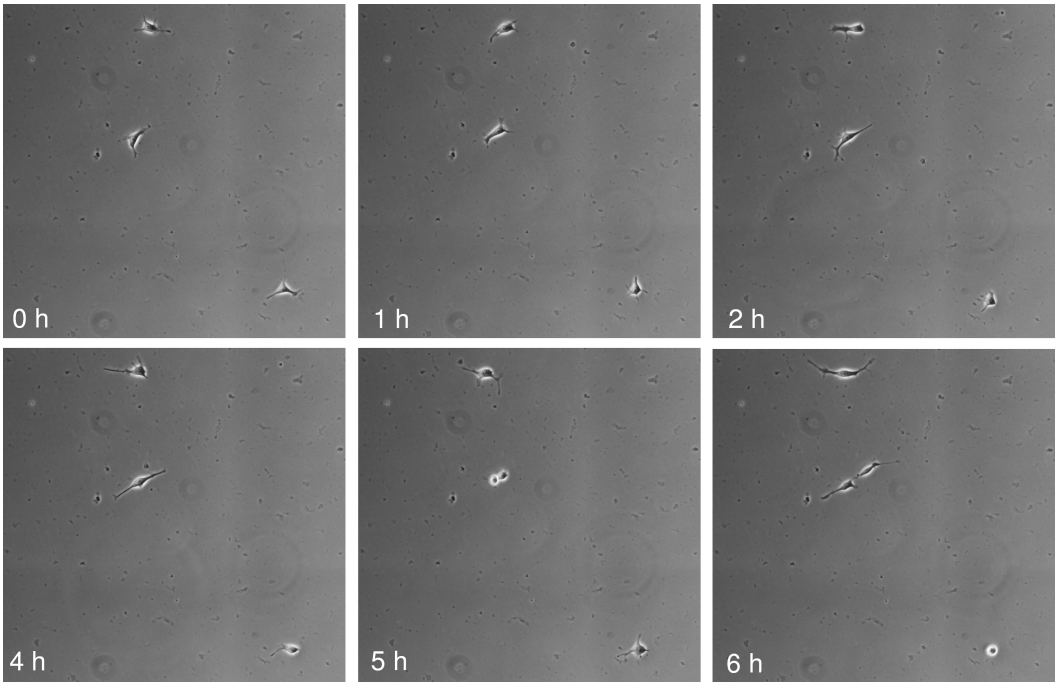

Figure 2.2: Example microscopy images from Experiment 1. Isolated cells were identified and their positions tracked.

2.2.1 Kinematic summary

In the absence (presence) of the CellTrace™, we identified 26 (22) initially isolated cells and tracked their movements manually. The key quantities extracted from the data are listed in Table 2.1. For cells in the absence of CellTrace™, of the 26 single cells, 24 divided, with this first division occurring, on average, at 7 hours. The remaining two cells died. In the presence of CellTrace™, 17 out of 22 single cells survived, with an average time to first division of 10 hours, considerably longer than for the cells without CellTrace™. Histograms of the time to the first division, t_d, for both groups are shown in Figure 2.3.

Parameter	Notation	Without CellTracer		With CellTracer	
		Mean	Median	Mean	Median
Number of cells	N	26		22	
% dead at 10 h		0%		23%	
% dead at 20 h		4%		23%	
Diffusivity (μm^2/h)	D	79.8 ± 5.2		20.5 ± 1.3	
Migration speed (μm/h)	v	22.7 ± 17.3 (0.6)	18.1 [10.2 31.8]	14.1 ± 12.2 (0.4)	10.9 [5.1 18.2]
Step length in x (μm)	l_x	3.5 ± 3.6 (0.1)	2.5 [1.2 5.0]	2.3 ± 2.5 (0.1)	1.9 [1.0 2.9]
Step length in y (μm)	l_y	3.5 ± 3.5 (0.1)	2.5 [1.2 5.0]	2.3 ± 2.6 (0.1)	1.9 [1.0 2.9]
Time to first division (h)	t_d	7 ± 5.2 (1.1)	5 [3.3 10.0]	10 ± 6.7 (1.6)	9 [5.9 12.6]
Correlation time (h)	τ	2.0±4.4 (0.2)	0.6 [0.2 1.5]	0.8±1.0 (0.04)	0.3 [0.1 0.8]

Table 2.1: Summary of parameters acquired for single cells. For means the errors are given as ± standard deviation (standard error) and for medians as [lower quartile upper quartile]. Note the diffusivity is a mean and 95% confidence interval by its definition. The step lengths and migration speed are averaged over all cells at all times (15 min intervals). The diffusivity was obtained using the fits to the MSD shown in Figure 2.9 and the correlation time from $\tau = 2D/v^2$ for instantaneous speeds.

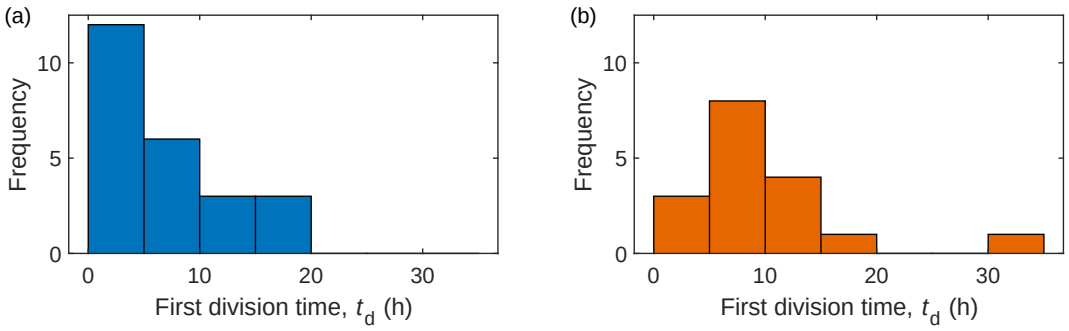

Figure 2.3: Histogram of division times, t_d with bin widths of 5 hours, for cells (a) without and (b) with CellTrace™.

Lineage trees to enable single cell identification are shown in Figure 2.4 and 2.5, in which cells are ordered according to their time to the first division (or death) and given an identification number. To illustrate the range of cell behaviours in both groups, four cells are highlighted and their trajectories and square displacements over time, L^2, are shown in Figure 2.4, 2.5 and 2.6. There are cells whose trajectories appear more tortuous (e.g., CellTrace™ cell 21) and others that exhibit directed motion (e.g., unstained cell 23, which travels a particularly long distance in the latter half of its cell cycle).

The cells can, and often do, change their direction of motion by up to π. An example is cell 22 (unstained) shown in Figure 2.7. The cell moves in the direction of its persistent pseudopodia protrusion, before contracting and moving in the direction of a new pseudopodia, resulting in a change of direction by approximately π. The whole manoeuvre in this example takes around 6 hours. Images of cell 22 during its migration are shown in Figure 2.7(a) with its full trajectory shown in Figure 2.7(b). This cell is elongated in the instantaneous direction of motion, with a pseudopodia protrusion leading its next movement. Across all cells the single cell shape can vary between approximately circular, with diameter of around $20\,\mu$m, to more elongated with length of up to $70\,\mu$m. In comparison, hESCs in colonies tend to be circular and considerably smaller, with diameters typically about $15\,\mu$m [87, 104, 188].

The average migration speeds are shown in Figure 2.8; notably, the median speeds of all cells are the same within errors within each staining group. For the unstained cells, the median instantaneous speed is $\overline{v} = 18.1\,\mu$m/h. This is significantly reduced for the stained cells, with $\overline{v} = 10.9\,\mu$m/h. The distribution of all instantaneous migration speeds is shown in Figure 2.8, showing a systematic skew towards intermediate speeds, and a significant reduction for the CellTrace™ cells. Note that the cells in Figure 2.8 are ordered according to the time to division/death as in Figure 2.4 and 2.5. The lack of any noticeable trend in the migration speed across the cells thus ordered indicates that there is no noticeable correlation between the cell migration speed and the time to division or death.

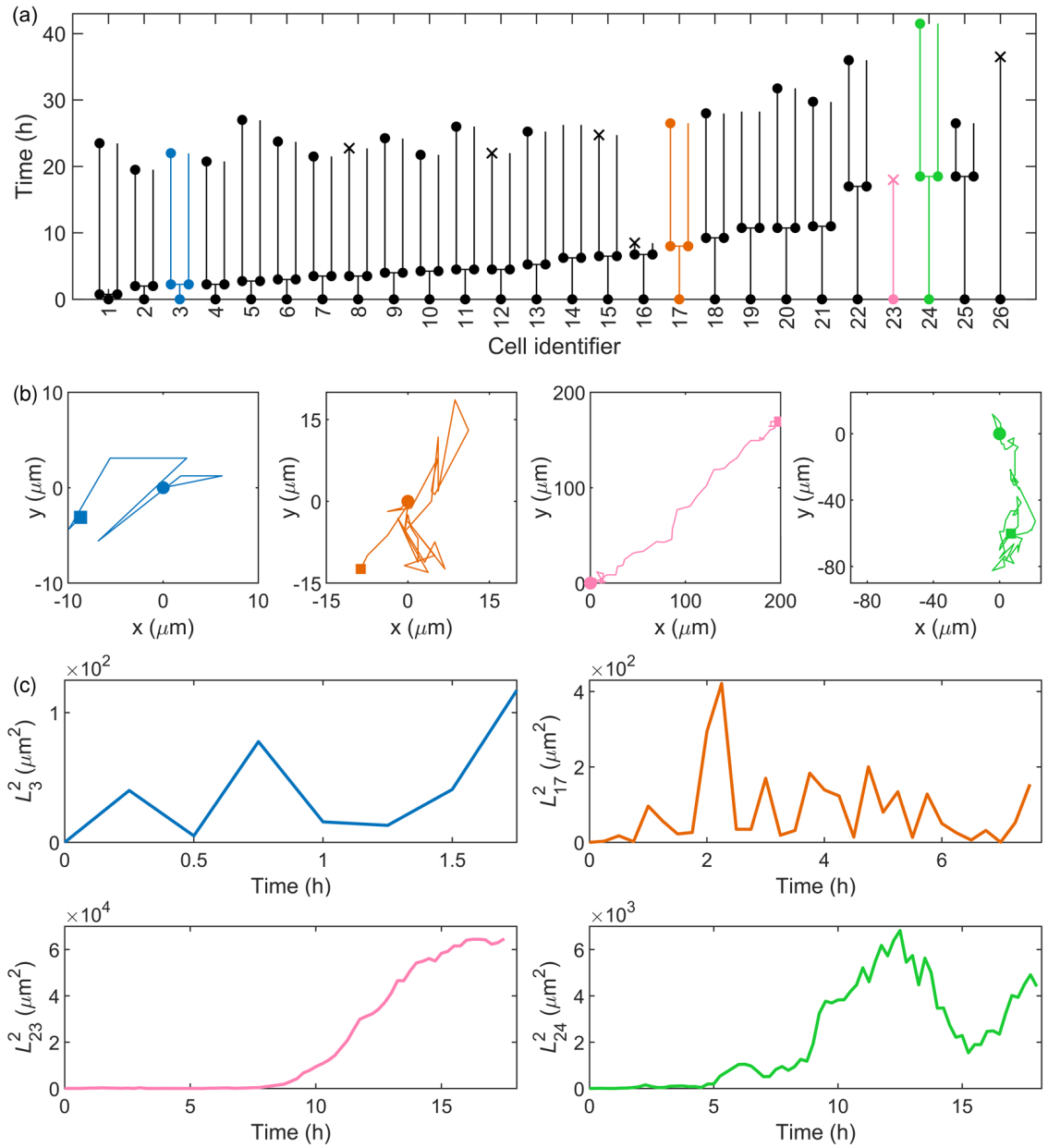

Figure 2.4: (a) The timelines of the 26 unstained single cells. A filled circle indicates that the cell has divided, a cross indicates the cell has died, and no marker indicates that the image was not clear enough to identify the cell confidently; in those cases, cells were not tracked beyond this point. (b) The corresponding trajectories of the cells highlighted in (a) with a circle and square signifying the start and end of the walk respectively. (c) L_i^2 as a function of time for the highlighted cells.

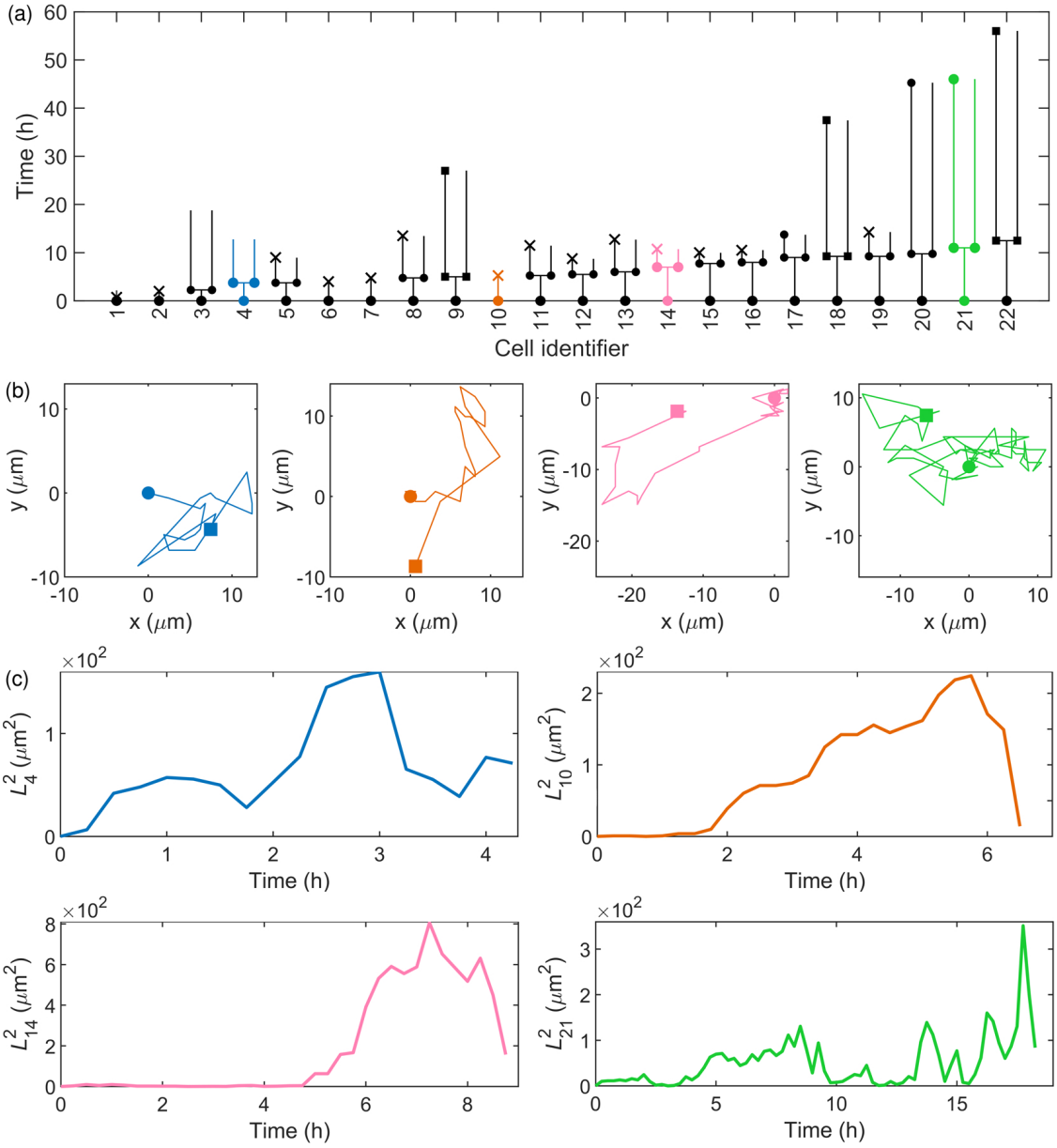

Figure 2.5: (a) The timelines of the 22 single cells stained with CellTrace™. A filled circle indicates that the cell has divided, a cross indicates the cell has died, a square indicates the cell joined a larger colony, and no marker indicates that the image was not clear enough to identify the cell confidently; in these cases, cells were not tracked beyond this point. (b) The corresponding trajectories of the cells highlighted in (a) with a circle and square signifying the start and end of the walk respectively. (c) L_i^2 as a function of time for the highlighted cells.

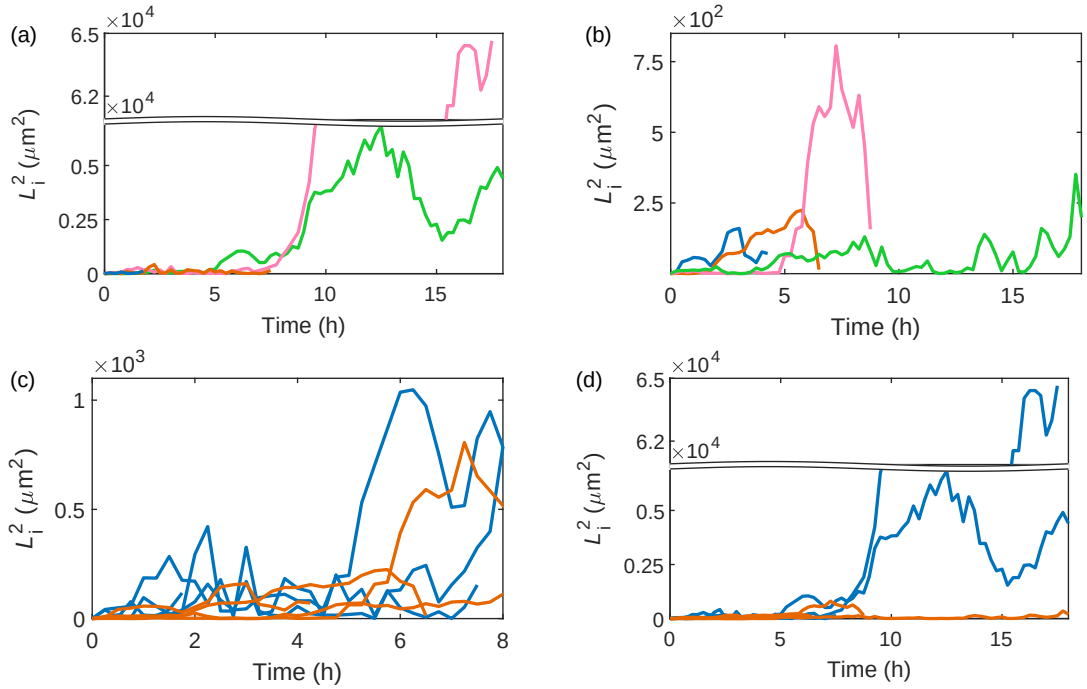

Figure 2.6: A comparison of L_i^2 over time for the highlighted cells in (a) Figure 2.4 (without staining) and (b) Figure 2.5 (with CellTrace™ staining). The same unstained (blue) and stained (orange) cells are shown on the same axes for up to (c) 8 hours and (d) 18 hours.

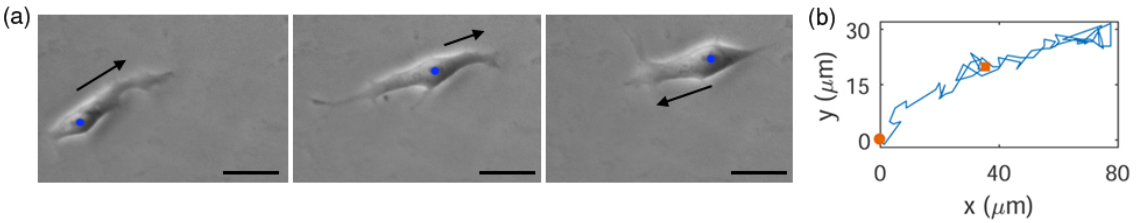

Figure 2.7: (a) Images of a migrating single hESC, unstained cell 22. The frames are taken at 15 min, 6 h 45 min and 14 h 15 min. The blue dot shows the cell nucleus and the black arrow the direction of instantaneous velocity. The scale bars are 30 μm in length. (b) Trajectory of the cell with the initial position (circle) and final position (square) shown.

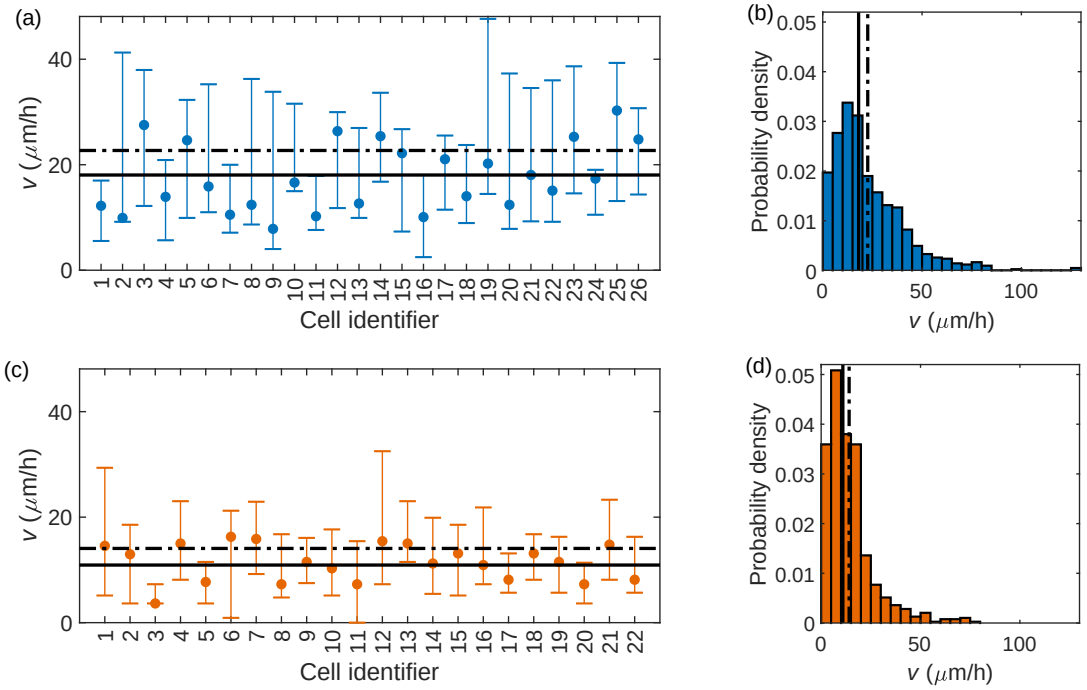

Figure 2.8: (a) The median migration speed of the unstained cells. The error bars correspond to the upper and lower quartiles. (b) Normalised histogram (probability density) of instantaneous migration speeds for all unstained cells. In (a) and (b) the median and mean migration speeds across all unstained cells, 18.1 [10.2 31.8] μm/h and $22.6 \pm 17.3\,(0.60)$ μm/h are shown by the solid and dashed lines, respectively. (c) The median migration speed of the cells with CellTrace™. (d) Normalised histogram of the instantaneous migration speed for all stained cells. In (c) and (d) the median and mean migration speeds across all stained cells, 10.9 [5.1 18.2] μm/h and $14.1 \pm 12.2\,(0.44)$ μm/h are shown by the solid and dashed lines, respectively.

2.2.2 Random walk analysis

Firstly, to verify if the cell migration is consistent, on average, with an isotropic random walk we consider the mean-square displacement, MSD= $\overline{L^2} = \overline{(\mathbf{x}(t) - \mathbf{x}_0)^2}$, averaged over the all cells within each staining category, shown in Figure 2.9. For unstained cells, the MSD is linear with time for the first seven hours of cell lifetimes with MSD= $(159.6 \pm 10.4)t$, giving the estimate of diffusivity as $D = 79.8 \pm 5.2\,\mu\text{m}^2/\text{h}$. Note that the number of samples (single cells) decreases over time due to cell death and division so the errors on the MSD increase over time. For the stained cells, the MSD is linear for the first eleven hours, with MSD= $(41.0 \pm 2.5)t$, giving $D = 20.5 \pm 1.3\,\mu\text{m}^2/\text{h}$. This further illustrates the negative effect CellTracer™ has on cell motility, with the diffusivity at early times a quarter that of the unstained group.

We can use these calculated diffusivities along with the cell velocities to estimate the correlation time, τ, of the walks using the relation $\tau = 2D/v^2$. Note that a value of τ is calculated for every cell, using the shared diffusivity D, but different instantaneous velocities. The average correlation times are, $\overline{\tau}$ =0.57 [0.22 1.53] h and $\overline{\tau}$ =0.31 [0.12 0.77] h, for the unstained and stained cells, respectively.

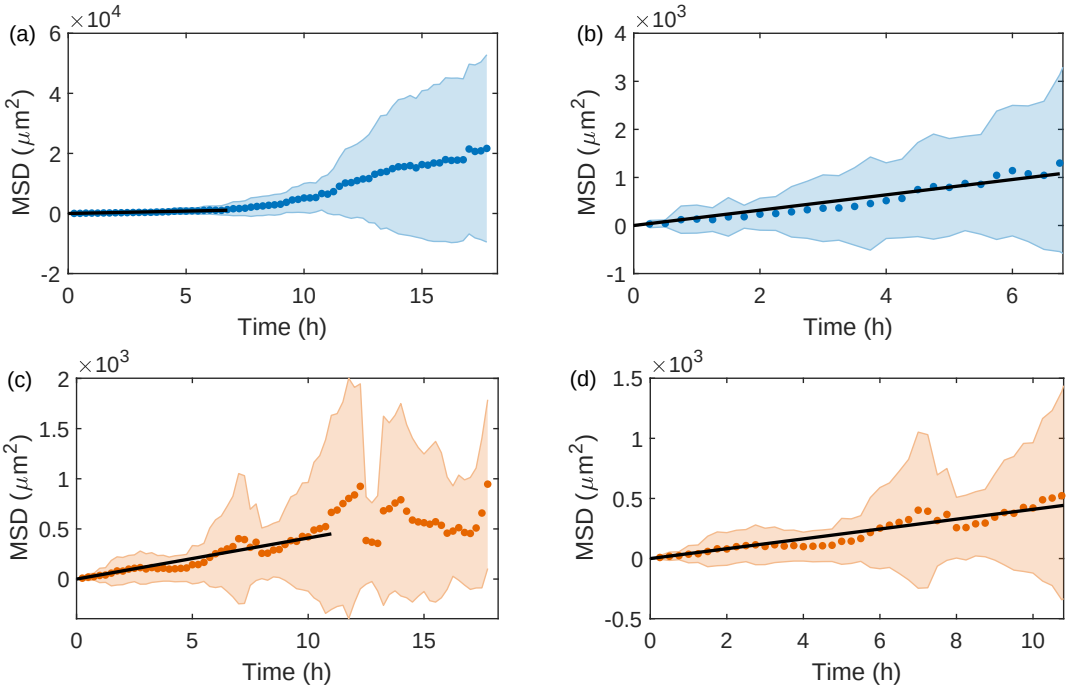

Figure 2.9: The MSD for unstained cells between (a) 0 and 17 hours and (b) 0 and 7 hours. The fit line, MSD= $(159.6 \pm 10.4)t$, $R^2 = 0.9$, is shown in black and the error bars are the standard deviation. The MSD for CellTrace™ stained cells between (c) 0 and 17 hours and (d) 0 and 11 hours. The fit line, MSD= $(41.0 \pm 2.5)t$, $R^2 = 0.9$, is shown in black and the error bars are the standard deviation.

A further method of characterising the tortuosity is the directionality, $\Delta = r/d$, with r displacement (the shortest distance between the cell's current position and its initial position), and d the total distance traversed. The average directionality for both groups with time is shown in Figure 2.10. Both show a quick reduction within the first four hours, but as expected, the directionality for the stained group remains lower through later times.

We can also evaluate the isotropic nature of the cell migration through the steps, δx and δy in the two orthogonal directions, x and y, per 15 minute frame in the microscopy imaging. The distributions of δx and δy are shown in Figure 2.11 for both staining groups. There is strong scatter in the individual step lengths but a clear reduction to smaller steps is seen for the stained cells. There is no evidence that the distributions between the x and y directions are different, but there is a significant difference between the staining groups, confirmed by the Kolmogorov-Smirnov test at the 95% confidence level.

If we consider the absolute step lengths, $l_x = |\delta x|$ and $l_x = |\delta y|$, the average values $\overline{l_x}$ and $\overline{l_y}$, are well defined, and are the same in each direction within statistical error, $\overline{l_x} = \overline{l_y} = 2.5\ [1.2\ 5.0]\,\mu$m. The Pearson product-moment correlation coefficient of l_x and l_y is 0.22, confirming the steps in the x and y directions are uncorrelated. The stained cells show a reduction in average step length, with $\overline{l_x} = \overline{l_y} = 1.9\ [1.0\ 2.9]\,\mu$m. This confirms that on average there is no bias in terms of the step length with relation to the fixed directional x and y axes.

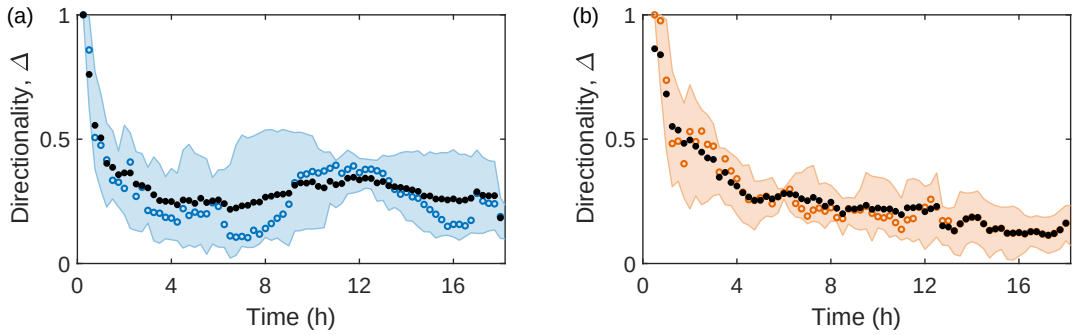

Figure 2.10: The directionality, $\Delta = r/d$ for (a) unstained and (b) stained cells. The blue and orange points show the median, with black points the mean and error bars the upper and lower quartiles.

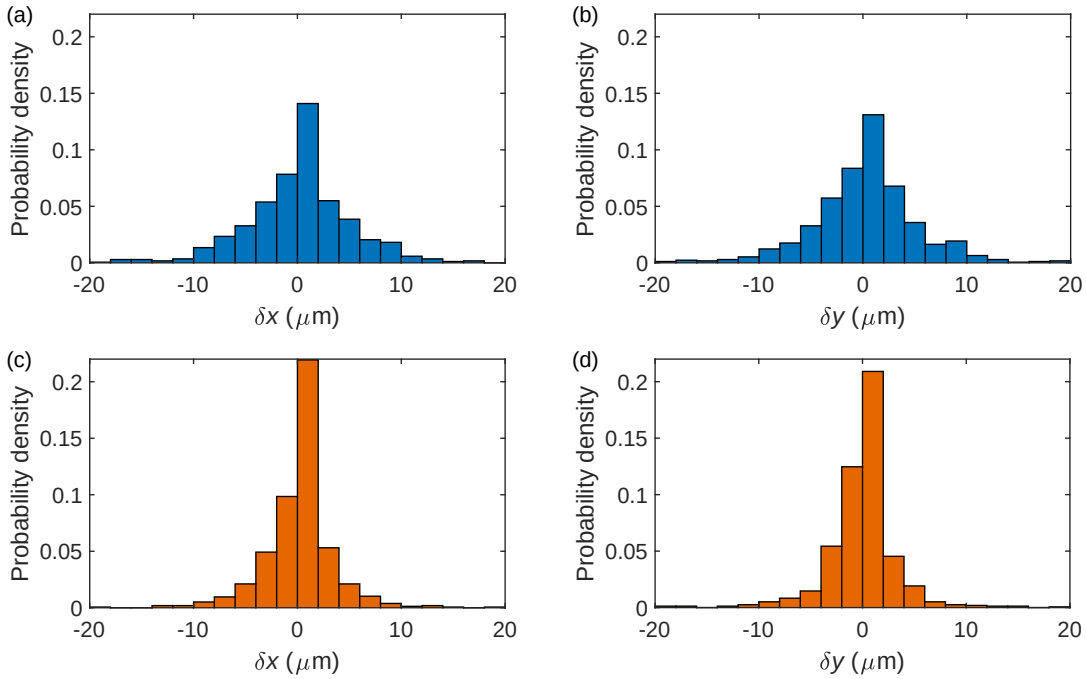

Figure 2.11: The distribution of the steps in the x and y directions, δx and δy for (a,b) unstained and (c,d) stained cells.

2.2.3 Correlations

Although on average the migration of the cells can be viewed as an isotropic random walk at early times in the cell cycle, the super-diffusive MSD (Figure 2.9) and the particular behaviour of cells such as unstained cell 22 (exhibiting directed motion before a rapid change of direction, Figure 2.7) and unstained cell 23 (directed at later times of the walk, Figure 2.4) suggests there might be more subtle aspects to the migration patterns. In this section we consider the correlations within the single cell movements (correlated random walks are described in Section 1.3.1). Since we have observed that the staining with CellTrace™ negatively impacts the motility of the cells (summarised in Table 2.1), here we include only the 26 unstained cells.

We can test for any correlations in the direction of the single cell movements by measuring the turning angle; the change in direction of the cell from one time frame to the next, denoted θ and illustrated in Figure 2.12(a). The polar histogram of θ shows the directionality of the turning angles for the 26 single cells, Figure 2.12(b). The corresponding linear histogram is shown in Figure 2.12(c). It is evident that the distribution has maxima at $\theta = 0$ and $\theta = \pi$: the cell preferentially moves directly forwards (i.e., continues in the same direction as the previous) or directly backwards (next step is at π from the previous), with a roughly equal frequency between the two directions. The bias is robust, remaining even if small steps ($< 7\,\mu$m) are removed from the dataset. The mean axis of movement (calculated using circular statistics described in Appendix B), shown in Figure 2.12(b), is approximately along the $\theta = 0$ or $\theta = \pi$ axis (with the standard deviation of $\sigma_\theta = 0.19$). In this manner, the motion represents a quasi-one-dimensional random walk.

The linear probability density distribution for θ, Figure 2.12(c), can be approximated by $a + b\cos(2\theta)$ with the parameters $a = 0.16 \pm 0.01$, $b = 0.04 \pm 0.02$ (with errors representing the 95% confidence interval) and an R^2 value of 0.6. Both the Rayleigh and V tests (for circular uniformity) [205, 206] reject the null hypothesis that the probability density of the turning angle θ is uniform at the 99.5% confidence level. Although containing some noise, this trigonometric fit suggests a symmetric spread of the distribution about the peaks $\theta = 0$ and $\theta = \pi$. This shows that the choice of θ varies in a periodic fashion, with a preference to move either in the same, or opposite direction as the previous movement.

The cell shown in Figure 2.7 is moving in this way, backwards and forwards along a preferential axis which is approximately aligned with its pseudopodia. We explore further how this correlated movement relates to the cell morphology in Section 2.2.4.

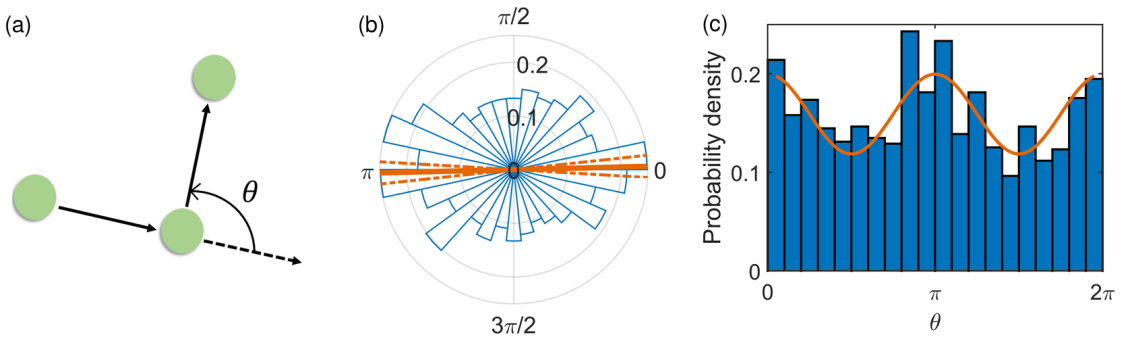

Figure 2.12: (a) The definition of the turning angle θ, the change in the cell's direction of motion from one time frame to the next. Green dots illustrate the positions of the cell in consecutive images with arrows representing the displacement vectors. (b) Polar histogram of θ for the 26 single cells, with 30 angular bins and 829 measurements. The mean value of 0.026, and one standard deviation 0.19 are shown as solid and dashed orange lines, respectively. (c) The probability density of θ binned into 20 intervals. The least-squares fit $0.16+0.04\cos(2\theta)$ is shown in orange.

The distribution of the turning angles has a distinct temporal pattern. The polar histograms of θ at early (0–5 h), intermediate (5–10 h) and late times (10–18 h) are shown in Figure 2.13. At early times, the distribution is slightly biased towards $\theta = \pi$, indicating a weak dominance of the back-and-forth motion over a systematic forward motion. However, this effect is weak and the distribution is approximately uniform over angles. This is consistent with our observations in Section 2.2.2 that the motion is close to an isotropic random walk at early times. By late times, however, the distribution is strongly biased towards $\theta = 0$, that is, persistent forward motion. On average for all times there is a mixture of persistent and back-and-forth motions.

This feature can be characterised with the temporal autocorrelation function, $C_\theta(\tau)$, for two-hourly moving averages of the angle θ. For each cell, $C_\theta(\tau)$ is calculated as the circular correlation for θ with itself, delayed by a time lag of τ. The average autocorrelation over all single cells, $\overline{C_\theta(\tau)}$, with least-squares fitting $\overline{C_\theta(\tau)} = e^{-\tau/\tau_c}$, $\tau_c = 0.8 \pm 0.1$ (95% confidence interval error), is shown in Figure 2.14(a). We see a temporal correlation in θ, with an average correlation decay time of $\tau_c = 0.8$ h.

We can also consider the angle between the cell displacement and the global frame (which does not change with time), denoted ϕ. There is no significant correlation in the global direction of movement, ϕ, when individual steps are considered. We attribute this to the dominance of the back-and-forth motion over short periods of time. However, considering two-hourly moving averages we again find a systematic trend. The average autocorrelation over all single cells, $\overline{C_\phi(\tau)}$, is shown in Figure 2.14(b). We see a temporal correlation in ϕ, with least-squares fitting $\overline{C_\phi(\tau)} = e^{-\tau/\tau_c}$, $\tau_c = 0.7 \pm 0.2$ shown in Figure 2.14(b) and hence a correlation decay time of $\tau_c = 0.7$ h, similar to that found

considering the change in direction θ. Notably there is significant anti-correlation (or anti-persistence) in ϕ for the time interval $[2\,\text{h} < \delta t < 5\,\text{h}]$, in agreement with the biased nature of the random walk.

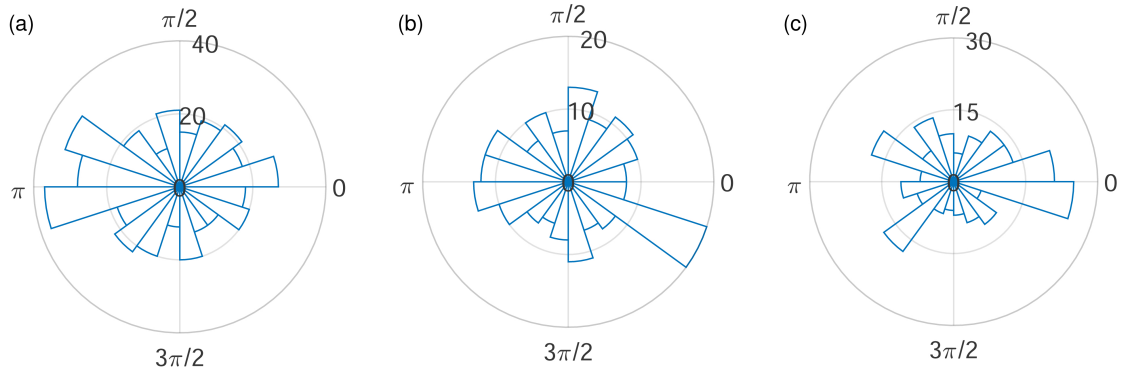

Figure 2.13: Polar histograms for θ for all 26 single cells in the time intervals (a) 0–5 h (26–12 cells, 20 bins and 404 measurements), (b) 5–10 h (12–8 cells, 20 bins and 197 measurements) and (c) 10–18 h (8–3 cells, 20 bins and 228 measurements). There are fewer cells at later times due to cell divisions and deaths.

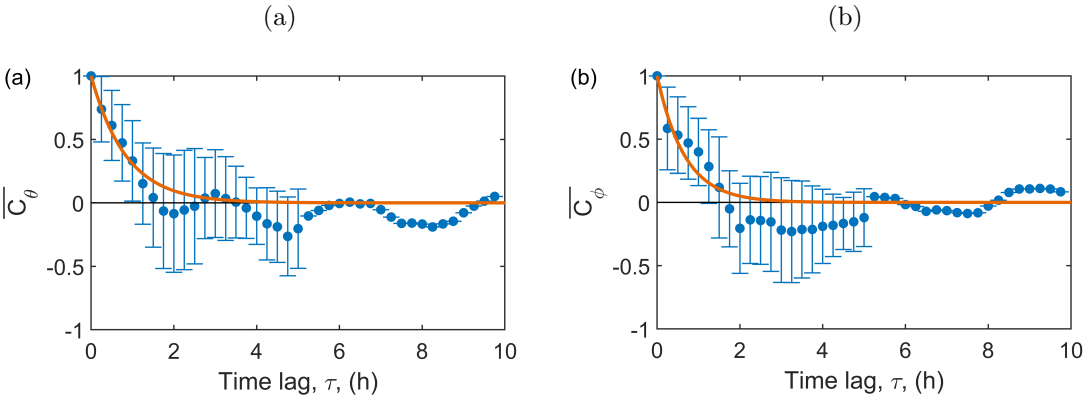

Figure 2.14: Average autocorrelation for (a) $\overline{C_\theta(\tau)}$ of θ and (b) $\overline{C_\phi(\tau)}$ of ϕ for single cells, with standard deviation error bars. The least squares fit (red line) is $\overline{C(\tau)} = e^{-\tau/\tau_c}$ with (a) $\tau_c = 0.8 \pm 0.1$ h and (b) $\tau_c = 0.7 \pm 0.2$ h for θ and ϕ, respectively. The autocorrelation for each cell is calculated for one third of their lifetimes and then averaged over all cells. Note that onwards from a time lag of 5 h, there is only one cell observed, hence the lack of error bars. Each lag corresponds to a time frame (15 min).

2.2.4 Movement and cell morphology

It is evident, from the images of cell 22 in Figure 2.7 in particular, that the direction of motion appears to be aligned with the elongation axis of the cell structure including its pseudopodia. This is unsurprising as cell branching and elongation has been shown to be involved in cell motion and directional persistence, although it has not been fully quantified [207]. A further example (cell 23) is shown in Figure 2.15(a). In this section we consider the trajectory of this directed single cell.

To analyse quantitatively the alignment of the direction of motion and the elongation of the cell, we measure the alignment angle of the cell, α, with respect to a global reference frame. Consider $\mathbf{R}(\alpha)$, the vector from the geometric centre to the boundary of the cell and $R_{\max}$ corresponding to the maximum magnitude of $\mathbf{R}$. The alignment angle α is defined as the angle between $R_{\max}$ and the horizontal, as shown in Figure 2.15(b). The polar histograms of α and the direction of travel on the plate, ϕ, both in the same global reference frame, are shown in Figure 2.15(c). Their mean values with 95% confidence interval errors are $\overline{\alpha} = 0.79 \pm 0.34$ and $\overline{\phi} = 0.72 \pm 0.35$. The difference between the means is insignificant as the Watson-Williams test (the circular analogue of the two-sample t-test) [206, 208] provides no evidence to reject the null hypothesis that $\overline{\alpha}$ and $\overline{\phi}$ are from the same distribution at the 99% confidence level. This shows that on average, the direction of travel is along the axis of cell alignment.

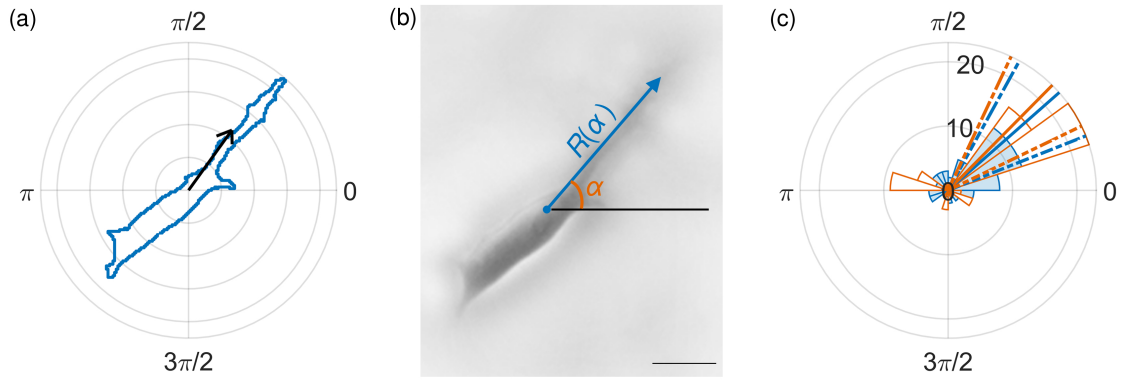

Figure 2.15: (a) The outline of the morphology of cell 23 (blue) which shows directed movement, with the geometric centre velocity (black arrow) at 11 h. (b) Corresponding microscopy image of cell 23 showing the geometric centre (blue dot), illustrating the definitions of the alignment angle α and the distance from the geometric centre to the edge of the cell $\mathbf{R}(\alpha)$. The scale bar is 30 μm. (c) Polar histogram of the alignment angle α (orange) and the direction of travel ϕ (blue, shaded) for cell 23 over a period of 17.5 h, with 70 measurements in 20 bins. The mean values are shown for ϕ (blue, 0.72 ± 0.35) and α (red, 0.79 ± 0.34) with the 95% confidence intervals for the mean shown as dashed lines.

We can also consider the temporal nature of the cell alignment and its direction of travel. Hourly moving averages of α and ϕ are shown in Figure 2.16(a), with a moderate correlation coefficient of 0.46. The speed of migration, v, and the measure of elongation of the cell, $R_{\max}/R_{\min}$, where $R_{\max} = \max|\mathbf{R}|$ and $R_{\min} = \min|\mathbf{R}|$, are shown as functions of time in Figure 2.16(b) and (c). If the cell is more elongated at a specific time point then $R_{\max} >> R_{\min}$ and so $R_{\max}/R_{\min}$ will be larger. Thus, high values of $R_{\max}/R_{\min}$ correspond to greater cell elongation and the minimum value of one corresponds to a circular shape. The hourly moving averages of $R_{\max}/R_{\min}$ and the cell speed v have a correlation coefficient of 0.53 suggesting a moderate positive correlation between the elongation of the cell and its speed. This suggests that directed movement is in the direction of the pseudopodia and that the cell moves faster when it is more elongated.

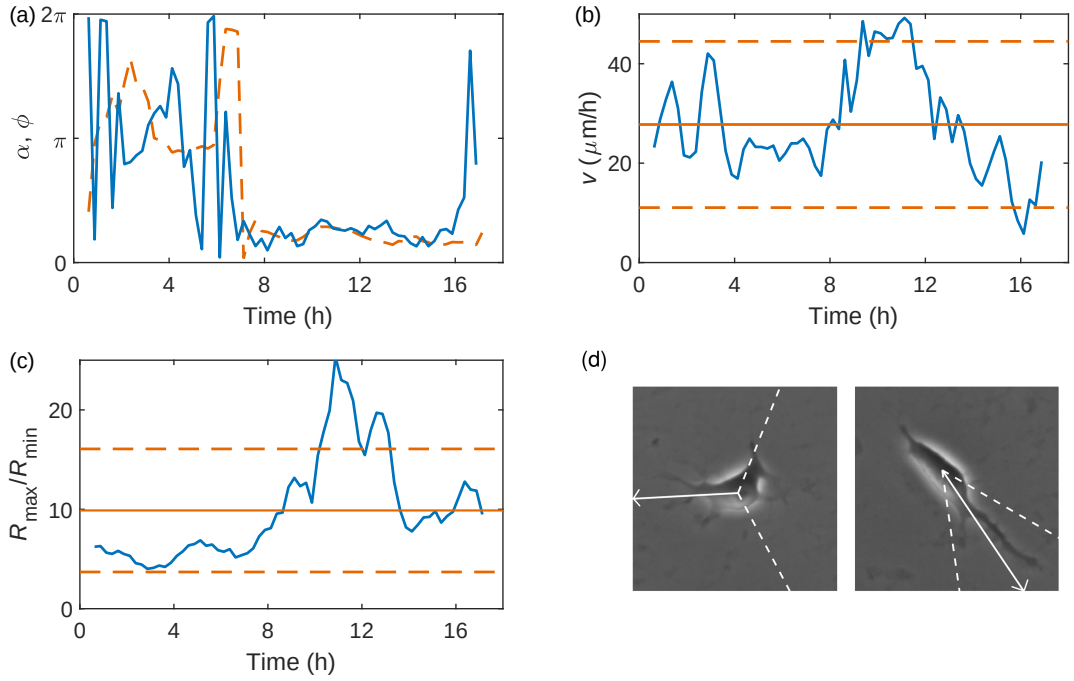

Figure 2.16: Hourly moving average of (a) ϕ (blue, solid) and α (orange, dashed) versus time, (b) the cell migration speed, v, and (c) $R_{\max}/R_{\min}$ over time. For (b) and (c) the solid lines show the mean values of $\overline{v} = 27.8\,\mu$m/h and $\overline{R_{\max}/R_{\min}} = 9.98$, with dashed lines one standard deviation from the mean ($\sigma_v = 16.7$ and $\sigma_{R_{\max}/R_{\min}} = 6.2$). (d) The cell at 4.5 and 11.5 h with a white arrow indicating the two-hourly average direction of the velocity with white dashed lines ± one standard deviation.

2.3 Discussion and conclusions

The rigorous quantification and characterisation of hESCs is integral to understanding colony formation from a single cell. In a last few years, several groups have employed time-lapse analysis to study the behaviour of hESCs during colony formation [21, 188]. However, their attention was focused on multicellular colonies rather than early stages of their formation. Individual cell migration has implications for colony clonality, with anomalous sub-diffusive cells contributing to clonality loss [21, 87]. Here we have quantified the behaviours of individual hESCs within the mathematical random walk framework, as a first step in the characterisation of colony growth from a single cell.

While the random walk framework provides a simplification of the migratory system, it is still important to keep in mind the mechanics of cell movement. In culture, hESCs are anchorage-dependent: they adhere to the surface and sense external cues by extending pseudopodia. For directed movement in response to external factors, cells acquire a defined front-rear polarity extending a protrusive structure at the leading edge before subsequently moving the cell body, and retracting the trailing edge [53, 202, 209]. The integration of negative and positive chemical feedback loops accounts for the oscillatory behaviour of pseudopodia, i.e., cycles of protrusion and retraction which result in cell movement [190, 202]. Observations of single cell movement in two-dimensional cultures, in the absence of external cues, indicate a production of pseudopodia structures in random directions, a behaviour observed in other cell types [210].

We have quantified the key migratory properties of single, isolated hESCs in the presence and absence of CellTrace™, a common biological stain for tracking cell generations. From these simple measures, there is evidence that the staining has a negative effect on cell motility, with cells moving less, and more slowly. This indicates that the application of such stains should be carefully considered, especially for clonal or low cell density assays in which they are most likely to have a negative impact.

We find that the average migration behaviour of isolated cells is diffusive, akin to an isotropic random walk up the first seven hours of evolution (with a reduced diffusivity for stained cells). These deviations after seven hours could be attributed to kinematic changes as cells begin the division process. Individual trajectories may show significant deviation from the average isotropic random walk behaviour, such as sporadic directed motion, but this is consistent with the probabilistic nature of individual behaviours. The typical cell displacements are up to several times the typical cell diameter, which is comparable to those considered in other studies of cell migration [87, 211].

In demonstrating the diffusive random-walk-like behaviour of the cells, our work opens the possibility to use the well-established mathematical theory of random walks and diffusion to help plan and optimise experiments with specific aims. For example,

agent-based models (described in Section 1.3.3), which combine the diffusive motion of cells with their biological state and interactions, have strong predictive power for mono-layer cultures, as demonstrated for epithelial cells [163].

The relative angle of movement, θ, characterises the dynamics of random walks further to the mean-square displacement [212]. Our results show that isolated single cells migrate in an unusual uni-directional walk, moving backwards and forwards along a preferred local axis, with cells becoming more persistent over time. Hence, the longest lived isolated cells show the strongest directional persistence. Broadly, there are a wide range of example cells that exhibit a preferential turning angle; those that can be modelled as a correlated random walk as previously discussed, e.g., [123–125]. There are also examples of a bi-modal preference for turning angle, similar to the one we see for single hESCs [213, 214]. Random walks with reversals are also seen in bacteria; a 'run-reverse' movement technique in which the cell moves in a directed manner before stopping, turning and travelling in the opposite direction [215–217]. A numerical analysis of a 2D random walker with non-uniform angular distribution is presented in [218], with an application to bacterial motion along a preferred direction. The bias in the walks of our cells is further shown in the temporal correlation in both the change in direction, and the direction of movement, with correlation decay times of around 0.8 h. The microscopy images in Figure 2.7 show the elongated morphology of the single cells, with movement in the direction of the leading pseudopodia, leading to this motion along a local axis.

These single cells demonstrate random migratory patterns, travel large distances and do not result in colony formation. Isolated cells seeded at low density display directional migration towards neighbours [188]. Perhaps in the absence of neighbours, as in this experiment, the cells employ the uni-directional walk along the local axis in an attempt to locate neighbours. It would be interesting to investigate with further experiments, in a similar manner to [188], how the presence of multiple neighbours, and their distances from the cell would affect this uni-directional behaviour. Our quantitative analysis of a directed cell trajectory confirms the axis of cell motion is aligned with the elongation axis of the cell. Increased elongation is also linked to increased speed, corresponding to previous results suggesting that persistence in direction of motion is linked to increased speed as a universal rule for all types of cells [219].

This analysis provides a quantification of the kinematic properties of isolated hESCs, which not only deepens our understanding of their characteristic movements, but also provides a basis for comparisons to future experiments under different conditions (for example, investigations into how the number of cells or cell seeding densities affect these migratory properties). This additional information on low density plated cells will assist in the development of agent-based models, combining the motion of diffusive and super-diffusive cells with their biological states and cell-cell interactions.

Chapter 3

Behaviour of pairs of hPSCs

Having characterised and quantified the kinematics of single hPSCs in Chapter 2 using experimental hESC data, we now extend the analysis to pairs of cells. In this Chapter we consider the dynamics of pairs of hESCs in close proximity, including their speeds, migration patterns and, in particular, the correlation between the individual motions of a pair's cells. The results we discuss here are also presented in [88] and [89].

3.1 Coordinated cell migration

The mechanics of individual cell migration for pairs of cells is the same as that for single cells, described in Section 2.1. The adaptations in cell morphology facilitate migration by the repeated cycle of pseudopodia extension and retraction [53, 202]. Although the individual cell mechanics remain the same, there are fundamental differences in the migratory behaviour of cells when there are neighbouring cells in close proximity.

Cells can communicate via inter-cellular (cell-cell) signalling which involves the secretion of molecules into the shared micro-environment (the substrate *in-vitro*) which then cause a response in other neighbouring cells [203, 220]. These signals (along with any other chemical gradients in the micro-environment) can stimulate chemotaxis: the movement of a cell in a direction corresponding to a gradient of increasing or decreasing concentration of a particular substance [203].

When cells get even closer together, they can adhere to one another through specialised molecules on their surface [221, 222]. This cell-cell adhesion suppresses individual cell migration but plays a crucial role in collective cell migration [223, 224], facilitating wound healing, cancer invasion, and embryonic development [225, 226]. It is therefore important to understand the changes in behaviour between adhering cells to characterise the mechanics of collective cell migration.

For hESCs, experimental results show cells can sense the presence of a neighbour up to $150\,\mu$m away. Particularly, if their separation distance is $< 70\,\mu$m, their movement becomes more organised and directional towards neighbouring cells [87, 188], suggesting that reduced separation distance between individual cells predisposes hESCs to migration toward one another, illustrated in Figure 3.1. This in turn has implications for cell clonality, with these cells contributing to migration induced clonality loss. In Chapter 2, for single cells we observed a directional walk along a preferential axis, aligned with cell elongation, which could indicate searching strategies for individual cells

to find neighbours, similar to other experimental results [87, 188].

The social environment of hPSCs also has implications for pluripotency, cell survival and colony proliferation, with an increase in neighbours correlating with increased pluripotency, survival rates and proliferation [188]. The dynamics of cells (not specific to stem cells) in a social context have been explored through an agent-based model which suggests that the diverse migratory behaviours are largely controlled by the changing environment [227].

It is clear that the effect of neighbours on the key properties of hPSCs, including cell migration, should be characterised to further our understanding of how they interact with the stem cell niche [228] and to facilitate the optimisation of *in-vitro* clonality. In the next section we extend our quantitative analysis of the migration of individual hESCs to pairs of cells to further explore the effect of cell neighbours on cell migration.

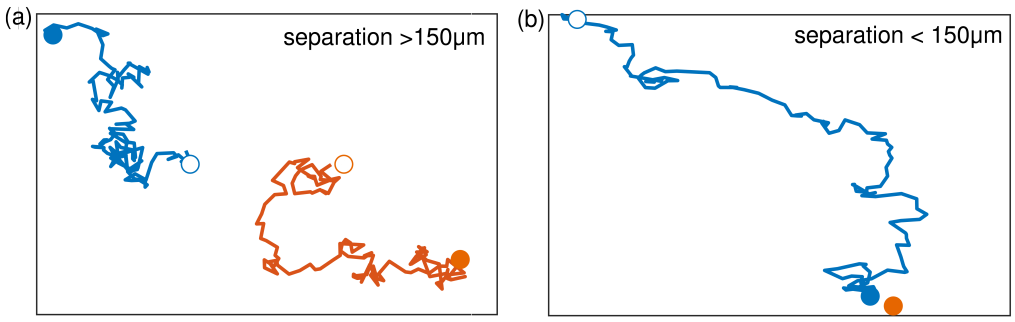

Figure 3.1: Illustration of experimental results [87, 188] which show that cells (a) $> 150\,\mu$m apart move as individual random walks while cells (b) $< 150\,\mu$m apart show more directed movement towards neighbours.

3.2 Pairs of hPSCs

In this section we quantify the migratory behaviours of pairs of hESCs in close proximity. We calculate their speeds, separations, and diffusivity. We consider the inherent correlations between pairs of cells and their movement patterns, and the implications cell mophology has on this movement. As in Chapter 2, we quantify the response of the cell dynamics to the CellTrace™ treatment, a common biological marker for tracing cell generations.

Experiment 1 The experiment is the same Experiment 1 presented in Chapter 2 and Appendix A tracking hESCs through time-lapse imaging. Upon cell division, the single hESCs split into a pair of genetically identical cells. These pairs, along with pairs seen together ($< 150\,\mu$m apart [87]) at the start of the recording (termed cells of unknown origin) were manually traced. Example microscopy images of the two types of pair are shown in Figure 3.2.

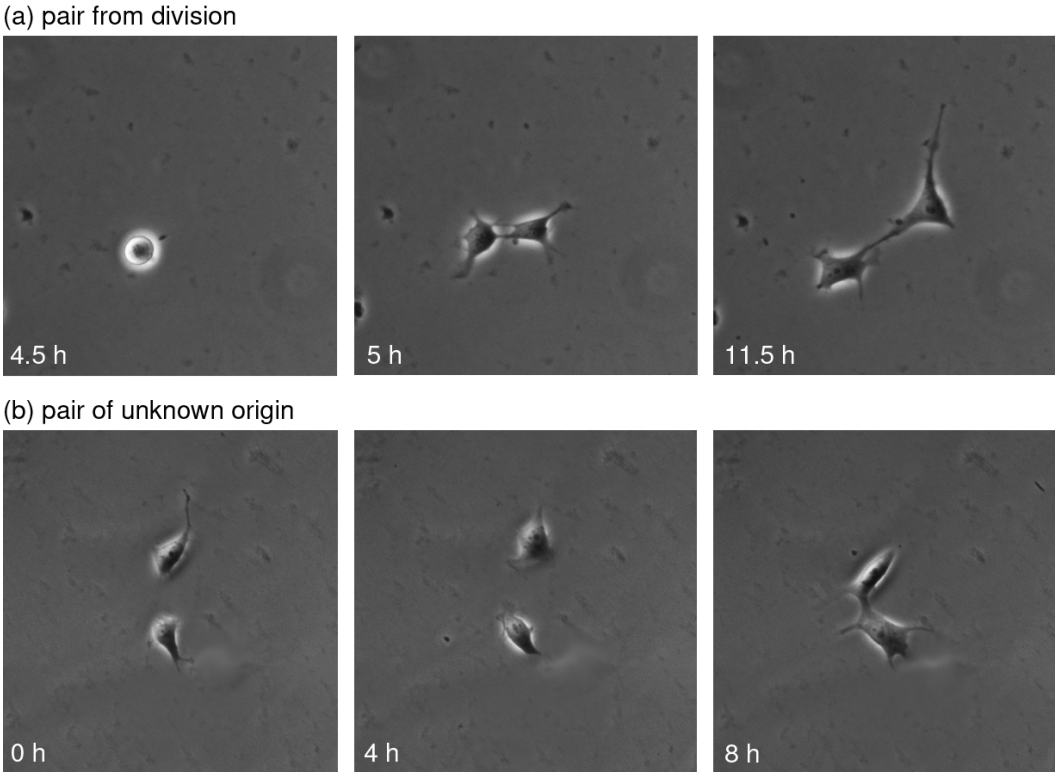

Figure 3.2: Example microscopy images from Experiment 1 for pair tracking for the two pair types: (a) pairs arising from a cell division and (b) pairs of unknown origin (seen together at the start of the microscopy recording).

3.2.1 Kinematic summary

To consider the impact of CellTrace™, we pool all cell pairs regardless of origin. In the absence (presence) of the CellTrace™, we identified 50 (18) pairs of cells. The key quantities extracted from the data are listed in Table 3.1.

The negative impact of the stain on cell survival (seen for single cells in Chapter 2) results in less traceable pairs in this group and higher rates of cell death, with 50% of pairs resulting in a cell death in the CellTrace™ group, compared to 14% in the unstained group. The stained pairs also show significantly reduced maximum and final separations, further indicative of their reduced motility.

Parameter	Notation	Without CellTracer		With CellTracer	
		Mean	Median	Mean	Median
Number of pairs	N	50		18	
% ending in death		14%		50%	
Diffusivity (μm^2/h)	D	83.2 ± 4.7		65.4 ± 5.5	
Migration speed (μm/h)	v	24.4 ± 19.9 (0.2)	19.4 [11.1 32.2]	20.4 ± 38.7 (1.0)	15.4 [8.6 22.6]
Step length in x (μm)	l_x	3.7 ± 3.8 (0.05)	2.5 [1.2 5.0]	2.9 ± 3.0 (0.08	1.9 [1.0 3.8]
Step length in y (μm)	l_y	4.0 ± 4.2 (0.05)	2.5 [1.2 5.6]	2.6 ± 2.7 (0.07)	1.9 [1.0 3.8]
Centroid speed (μm/h)	v_{pc}	18.3 ± 13.7 (0.2)	15.1 [8.9 23.9]	14.2 ± 11.1 (0.4)	11.7 [6.9 18.1]
Relative speed (μm/h)	v_{r}	33.6 ± 27.2 (0.5)	26.7 [15.1 43.9]	21.3 ± 17.0 (0.6)	17.2 [10.9 27.2]
Initial separation (μm)	s_0	35.6 ± 26.5 (3.7)	23.3 [16.5 47.3]	17.5 ± 6.2 (1.5)	15.9 [13.5 22.1]
Final separation (μm)	s_{f}	58.9 ± 60.3 (8.5)	35.7 [23.7 63.6]	19.4 ± 6.9 (1.6)	19.4 [14.7 23.7]
Max separation (μm)	s_{max}	106.7 ± 62.0 (8.8)	87.8 [67.2 130.9]	27.0 ± 8.4 (2.0)	26.3 [21.3 33.9]
Time to s_{max}	–	8.5 ± 6.5 (0.9)	7.0 [2.8 13.0]	6.9 ± 7.6 (1.8)	3.5 [1.3 8.8]
Correlation time (h)	τ	3.5 ± 11.2 (0.2)	0.7 [0.3 2.1]	3.3 ± 6.5 (0.3)	1.0 [0.4 2.7]

Table 3.1: Summary of parameters acquired for pairs of cells. For means the errors are given as ± standard deviation (standard error) and for medians as [lower quartile upper quartile]. Note the diffusivity is a mean and 95% confidence interval by its definition. The migration speeds are averaged over all cells at all times (15 min intervals). The diffusivity was obtained using the fits to the MSD shown in Figure 3.6 and the correlation time from $\tau = 2D/v^2$ for instantaneous centroid speeds.

In addition to each individual cell's instantaneous velocity, $\boldsymbol{v}$, the movement of a pair of cells can be characterised by two distinct velocities. The first of these is the velocity of the pair as a whole, i.e., its pair centroid velocity $\boldsymbol{v}_{\mathrm{pc}}$. Denoting the individual velocities of the cells in the pair as $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$, then the pair centroid velocity is the vectorial average of these, i.e., $\boldsymbol{v}_{\mathrm{pc}} = (\boldsymbol{v}_1 + \boldsymbol{v}_2)/2$. The second of these quantities is the relative velocity of the cells within the pair $\boldsymbol{v}_{\mathrm{r}}$, defined as the difference of the individual velocities, $\boldsymbol{v}_{\mathrm{r}} = \boldsymbol{v}_1 - \boldsymbol{v}_2$. This velocity characterises the approach of the cells or their motion away from each other.

From these two quantities we define the corresponding speeds as the vector magnitudes, $v = |\boldsymbol{v}|$, $v_{\mathrm{pc}} = |\boldsymbol{v}_{\mathrm{pc}}|$ and $v_{\mathrm{r}} = |\boldsymbol{v}_{\mathrm{r}}|$. The individual, centroid and relative speeds for all pairs are shown in Figure 3.3, 3.4 and 3.5.

The median individual pair speeds are 19 μm/h and 15 μm/h for the unstained and stained groups respectively, similar to that for the individual single cells (medians 18 μm/h and 11 μm/h) and showing a similar reduction in motility for the stained cells. The median pair centroid speeds, v_{pc}, and relative speeds, v_{r} are 15 μm/h (unstained) and 12 μm/h (stained), and 27 μm/h (unstained) and 17 μm/h (stained), showing that the pair as a whole moves at a similar speed to a single cell, but the cells move relatively fast within the pair.

Of the 50 unstained pairs, 23 pairs originated from divisions of the single cells discussed in Chapter 2. The remaining 27 pairs of unknown origin were seen together at the start of the image recording and so it is not clear if the pair formed from the division of a single cell or two unrelated single cells migrating towards each other. Although in the following sections the analysis is given for all 50 cells together, for completeness the corresponding kinematic parameters for the two types of pairs are given in Table 3.2. For the CellTrace™ pairs, 17 of the 18 cells originated from a cell division, so they cannot be considered in two separate groups.

Parameter	Notation	From division		Unknown origin	
		Mean	Median	Mean	Median
Number of pairs	N	23		27	
Diffusivity (μm^2/h)	D	66.3 ± 3.3		42.8 ± 2.2	
Migration speed (μm/h)	v	25.5 ± 19.2 (0.5)	20.5 [10.2 29.2]	23.2 ± 20.5 (0.4)	17.9 [10.2 30.0]
Step length in x (μm)	l_x	3.6 ± 3.9 (0.07)	2.5 [0.6 5.0]	3.4 ± 3.7 (0.06	2.5 [1.2 4.3]
Step length in y (μm)	l_y	3.6 ± 3.7 (0.06)	2.5 [0.6 5.9]	3.9 ± 4.4 (0.08)	2.5 [1.2 5.0]
Centroid speed (μm/h)	v_{pc}	18.6 ± 12.9 (0.3)	15.9 [9.0 24.8]	17.6 ± 14.1 (0.3)	14.1 [8.7 22.8]
Relative speed (μm/h)	v_{r}	35.2 ± 25.8 (0.6)	29.0 [17.5 45.5]	31.5 ± 28.3 (0.7)	24.5 [12.6 40.1]
Initial separation (μm)	s_0	18.1 ± 4.6 (0.9)	16.9 [14.4 21.2]	50.6 ± 28.3 (5.5)	46.5 [25.5 69.7]
Final separation (μm)	s_{f}	76.7 ± 73.2 (15.3)	47.6 [26.1 94.5]	43 ± 42.3 (8.1)	28.7 [22.8 28.7]
Max separation (μm)	s_{max}	124.3 ± 76.2 (15.9)	92.8 [78.7 171.0]	91.6 ± 42.7 (8.2)	87.6 [65.5 110.7]
Time to s_{max}	–	11.0 ± 6.4 (1.3)	10.3 [6.1 16.8]	6.4 ± 5.9 (1.1)	4.3 [2.3 10.3]
Correlation time (h)	τ	1.6 ± 5.0 (0.1)	0.3 [0.1 1.0]	3.1 ± 9.9 (0.2)	0.7 [0.3 1.7]

Table 3.2: Summary of parameters acquired for pairs of unstained cells. For means the errors are given as ±standard deviation (standard error) and for medians as [lower quartile upper quartile]. Note the diffusivity is a mean and 95% confidence interval by its definition. The migration speeds are averaged over all cells at all times (15 min intervals). The diffusivity was obtained using the fits to the MSD and the correlation time from $\tau = 2D/v^2$ for instantaneous centroid speeds.

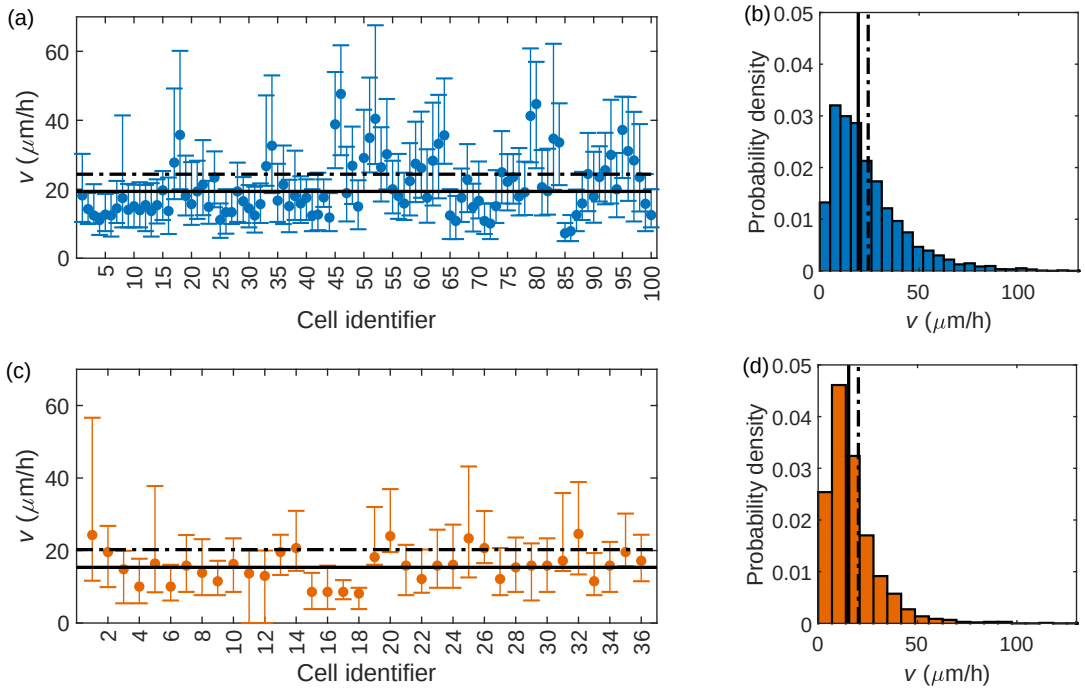

Figure 3.3: (a) The median migration speed of the unstained cell pairs. The error bars correspond to the upper and lower quartiles. (b) Normalised histogram (probability density) of instantaneous migration speeds for all unstained pairs of cells. In (a) and (b) the median and mean migration speeds across all unstained cells, $19.4\,[11.1\ 32.2]\,\mu$m/h and $24.4 \pm 19.9\,(0.2)\,\mu$m/h are shown by the solid and dashed lines, respectively. (c) The median migration speed of the cell pairs with CellTrace™. (d) Normalised histogram of the instantaneous migration speed for all stained cells. In (c) and (d) the median and mean migration speeds across all stained cells, $15.4\,[8.6\ 22.6]\,\mu$m/h and $20.4 \pm 38.7\,(1.0)\,\mu$m/h are shown by the solid and dashed lines, respectively.

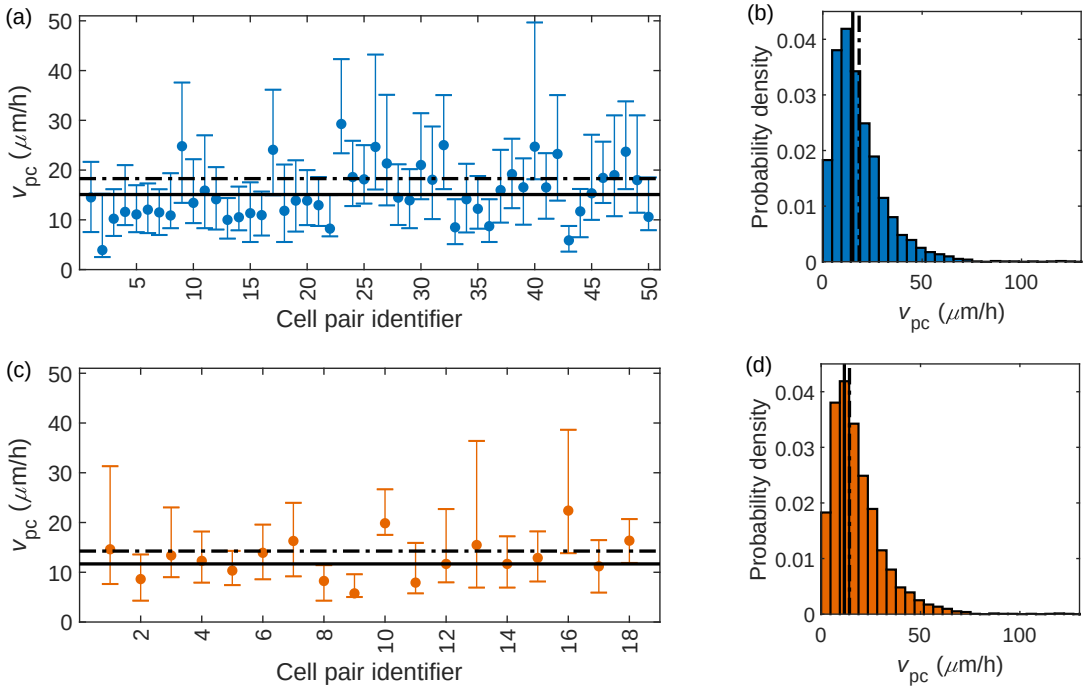

Figure 3.4: (a) The median pair centroid migration speeds for unstained pairs. The error bars correspond to the upper and lower quartiles. (b) Normalised histogram (probability density) of all instantaneous centroid speeds for all unstained pairs of cells. In (a) and (b) the median and mean migration speeds across all unstained cells, 15.1 [8.9 23.9] μm/h and 18.3 ± 13.7 (0.2) μm/h are shown by the solid and dashed lines, respectively. (c) The median pair centroid migration speed of the cell pairs with CellTrace™. (d) Normalised histogram of the instantaneous migration speed for all stained cells. In (c) and (d) the median and mean migration speeds across all stained cells, 11.7 [6.9 18.1] μm/h and $14.2 \pm$ 11.1 (0.4) μm/h are shown by the solid and dashed lines, respectively.

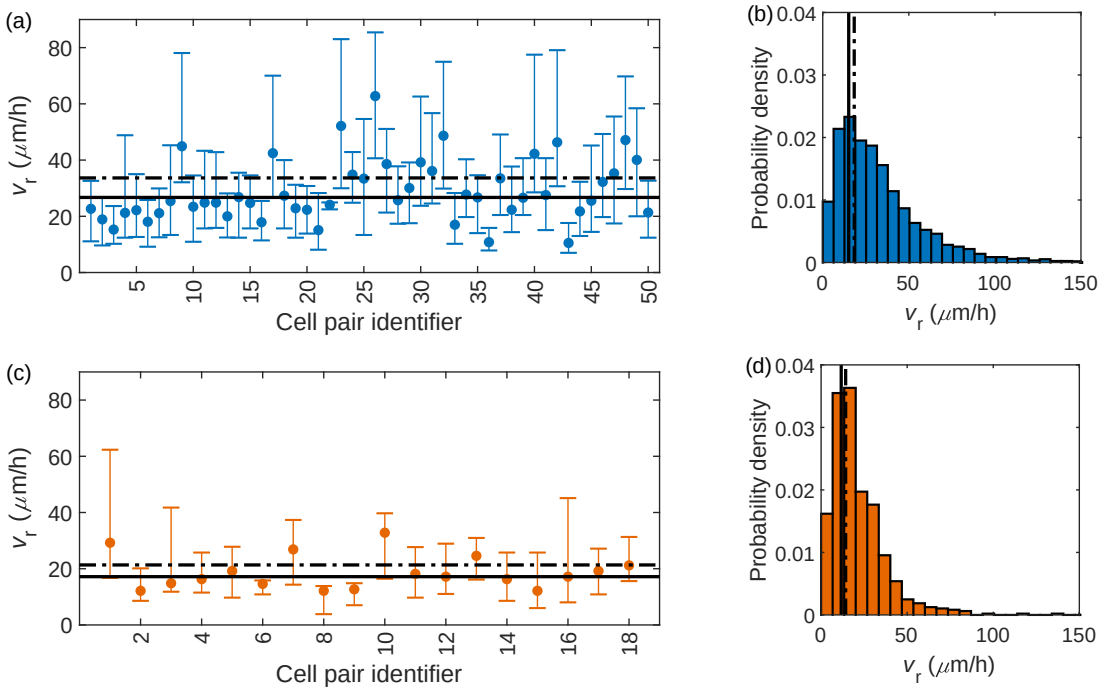

Figure 3.5: (a) The median relative migration speeds without CellTrace™. The error bars correspond to the upper and lower quartiles. (b) Normalised histogram (probability density) of all instantaneous relative speeds for all unstained pairs of cells. In (a) and (b) the median and mean migration speeds across all unstained cells, 26.7 [15.1 43.9] μm/h and 33.6±27.2 (0.5) μm/h are shown by the solid and dashed lines, respectively. (c) The median relative migration speed of the cell pairs with CellTrace™. (d) Normalised histogram of the instantaneous relative speeds for all stained cells. In (c) and (d) the median and mean migration speeds across all stained cells, 17.2 [10.9 27.2] μm/h and 21.3 ± 17.0 (0.6) μm/h are shown by the solid and dashed lines, respectively.

3.2.2 Random walk analysis

In this section we will consider the motion of pairs within the random walk framework described in Section 1.3.1. It can be expected that the relative velocity of the cells in a pair can no longer be approximated by an isotropic random walk because of their mutual interaction [87]. However, the pair as a whole still might perform a random walk, but perhaps with a different diffusivity from that of a single cell. In this case we can consider the mean-square displacement for the cell centroid position, MSD= $\overline{L^2} = \overline{(\mathbf{x}(t) - \mathbf{x}_0)^2}$, shown in Figure 3.6.

For unstained cells, the centroid MSD is linear with time for the first 12 hours of cell lifetimes with MSD= $(166.3 \pm 9.4)t$, giving the estimate of diffusivity as $D = 83.2 \pm 4.7\,\mu\text{m}^2/\text{h}$, similar to that of single cells. For the stained cells, the MSD is only linear for the first seven hours, with MSD= $(130.7 \pm 10.9)t$, giving $D = 65.4 \pm 5.5\,\mu\text{m}^2/\text{h}$. This further illustrates the negative effect CellTracer™ has on cell motility, with a 21% reduction in the diffusivity at early times compared to the unstained group. If the unstained pairs are split by origin, the pairs arising from a division have a higher diffusivity ($D = 66.3 \pm 3.3\,\mu\text{m}^2/\text{h}$) than those of unknown origin ($D = 42.8 \pm 2.2\,\mu\text{m}^2/\text{h}$). On average all the pairs show super-diffusive behaviour at later times. More experimental studies are required to clarify the significance and biological cause of this longer-term behaviour.

We can use these calculated diffusivities along with the cell centroid velocities to estimate the correlation time, τ, of the walks using the relation $\tau = 2D/v^2$. Note that a value of τ is calculated for every cell, using the shared diffusivity D, but different instantaneous velocities. The average correlation times are, $\overline{\tau}$ =0.7 [0.3 2.1] h and $\overline{\tau}$ =1.0 [0.4 2.7] h, for the unstained and stained cells, respectively.

A further method of characterising the tortuosity, is the directionality, $\Delta = r/d$, with r displacement (the shortest distance between the cell's current position and its initial position), and d the total distance traversed. The average directionality with time for the centroid position of both stained and unstained cells is shown in Figure 3.7. Both show a quick reduction within the first four hours, but as expected, the directionality for the stained group remains lower through later times.

We can also evaluate the isotropic nature of the individual cell migration within a pair through the steps, δx and δy in the two orthogonal directions, x and y, per 15 minute frame in the microscopy imaging. The distributions of δx and δy are shown in Figure 3.8 for both staining groups. There is strong scatter in the individual step lengths but a clear reduction to smaller steps is seen for the stained cells. There is no evidence that the distributions between the x and y directions are different, but there is a significant difference between the staining groups (confirmed by Kolmogorov-Smirnov test at the 95% confidence level).

If we consider the absolute step lengths, $l_x = |\delta x|$ and $l_x = |\delta y|$, the average values for

unstained cells within a pair, $\overline{l_x}$ and $\overline{l_y}$, are well defined, and are the same in each direction within statistical error, $\overline{l_x} = \overline{l_y} = 2.5$ [1.2 5.6] μm. The correlation coefficient of l_x and l_y is 0.22, confirming the steps in the x and y directions are uncorrelated. The stained cells show a reduction in average step length, with $\overline{l_x} = \overline{l_y} = 1.9$ [1.0 3.8] μm. This confirms that on average there is no bias in terms of the step length with relation to the fixed directional x and y axes. The step lengths are the same as for cells in isolation, indicating that the presence of a close neighbour has no effect on the step length.

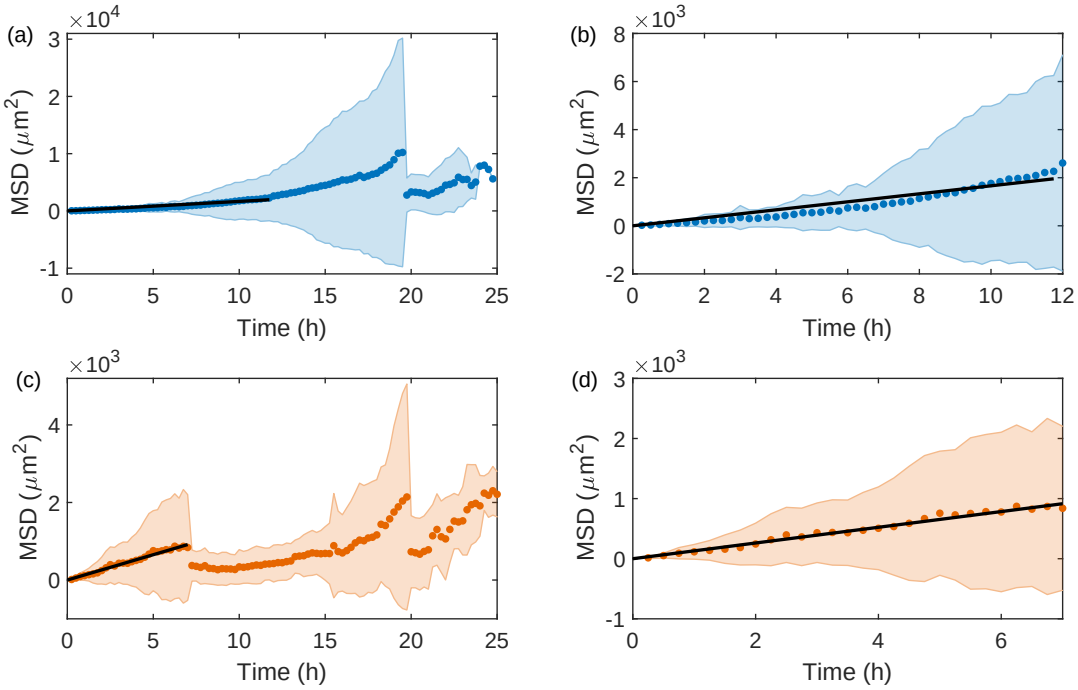

Figure 3.6: The MSD for the centroid position of unstained cells between (a) 0 and 25 hours and (b) 0 and 12 hours. The fit line, MSD= $(166.3 \pm 9.4)t$, $R^2 = 0.9$, is shown in black and the error bars are the standard deviation. The MSD for CellTrace™ stained cells between (c) 0 and 25 hours and (d) 0 and 7 hours. The fit line, MSD= $(130.7 \pm 10.9)t$, $R^2 = 0.8$, is shown in black and the error bars are the standard deviation.

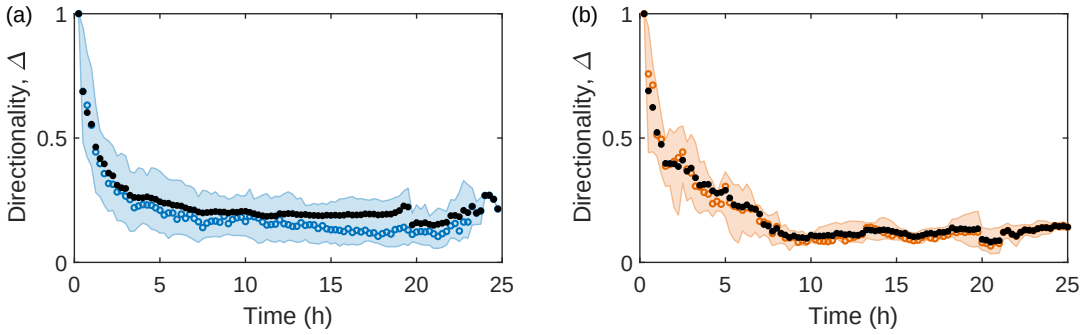

Figure 3.7: The directionality, $\Delta = r/d$ for (a) unstained and (b) stained pairs of cells. The blue and orange points show the median, with black points the mean and error bars the upper and lower quartiles.

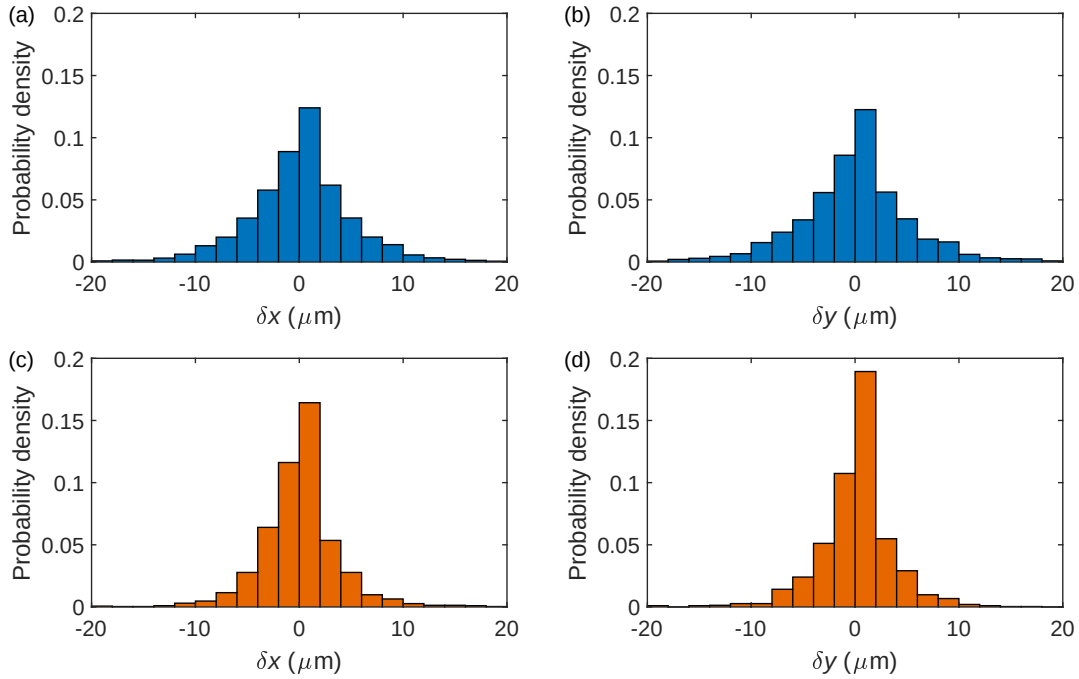

Figure 3.8: The distribution of the steps in the x and y directions, δx and δy for (a b) unstained and (c,d) stained cells within a pair.

3.2.3 Correlations

In the previous section we showed that the motion of the geometric centre of cell pairs can be approximated by an isotropic random walk for up to around 12 h of their evolution, with a similar diffusivity to single cells. In this section we look in greater detail at the dynamics of these pairs of hESCs, in particular the correlations between the individual motions of a pair's cells. We consider the 50 pairs from the unstained group, since we have found that CellTrace™ has a negative impact on cell motility and survival.

Firstly, we can consider the separation of all the pairs on average. The mean separation with time t, $\overline{r(t)}$, up to 18 hours is shown in Figure 3.9. By performing a least-squares fit of the functional form $\overline{r} = \mathrm{A} - \mathrm{B}e^{-t/\mathrm{C}}$, for parameters A, B and C we obtain the line $\overline{r} = (68 \pm 0.6) - (37 \pm 3)e^{-t/(2\pm0.03)}$. The asymptotic nature of $\overline{r}$ indicates an optimal separation of pairs at around 70 μm. This could be due to morphological pseudopodia connections between pairs, which we discuss further in Section 3.2.4.

We can consider the change in direction at each timestep, θ, for each individual cell within a pair, as illustrated in Figure 3.10(a). The polar histogram of θ shows the directionality of the turning angles for the 100 cells within a pair, Figure 3.10(b). The corresponding linear histogram is shown in Figure 3.10(c). It is evident that the distribution has maxima at $\theta = 0$ and $\theta = \pi$, similar to the single cells, however here the maxima at $\theta = \pi$ is dominant. This indicates that the individual cells within a pair preferentially move directly forwards (i.e., continue in the same direction as the previous) or directly backwards (next step is at π from the previous), with a change in direction more frequent than persistence. The mean axis of movement (calculated using circular statistics, Appendix B), shown in Fig. 3.10(b), is approximately along the $\theta = 0$ or $\theta = \pi$ axis (with the standard deviation of $\sigma_\theta = 0.05$). In this manner, the motion represents a quasi-one-dimensional random walk. The global maxima at $\theta = \pi$ means that the linear probability density distribution, Figure 3.10(c), can not be approximated by $a + b\cos(2\theta)$

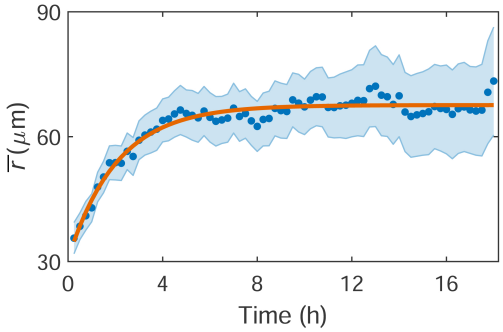

Figure 3.9: The mean separation, $\overline{r}$, for pairs over time with fit $\overline{r} = (68 \pm 0.6) - (37 \pm 3)e^{-t/(2\pm0.03)}$ (orange) and $R^2 = 0.94$. The error bars show the standard error.

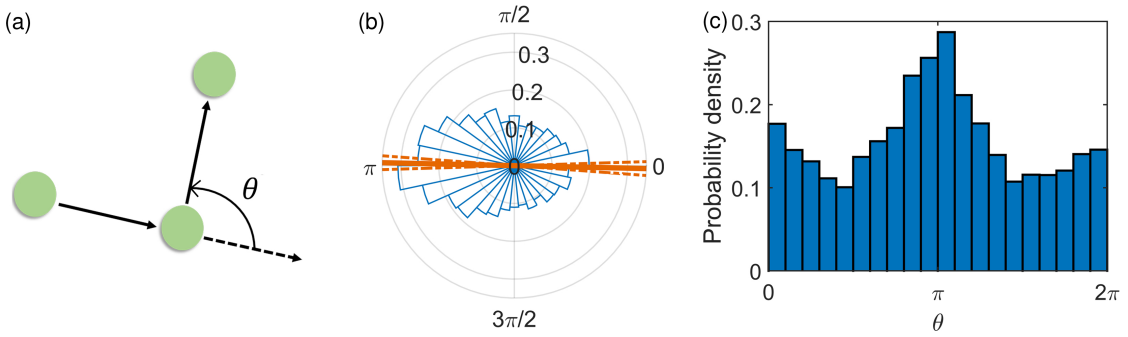

Figure 3.10: (a) The definition of the turning angle θ, the change in the cell's direction of motion from one time frame to the next. Green dots illustrate the positions of the cell in consecutive images with arrows representing the displacement vectors. (b) Polar histogram of θ for the 100 cells belonging to a pair, with 30 angular bins and 6548 observations. The mean value of 3.12, and one standard deviation 0.05 are shown as solid and dashed orange lines, respectively. (c) The corresponding linear histogram.

as for the single cells. However, the distribution still shows a roughly symmetric spread about the peaks $\theta = 0$ and $\theta = \pi$ and suggests the choice of θ varies in a periodic fashion, with a preference to move in the opposite direction to the previous movement.

To quantify the coordination between the movements of the two cells in a pair, we can measure the smaller angle between their velocities, $0 < \psi < \pi$, illustrated in Figure 3.11(a). If the cells travel in the same direction on the plate, then $\psi = 0$, and if they travel in opposite directions $\psi = \pi$: note that $\psi = \pi$ does not distinguish between the two cells moving exactly towards each other or exactly apart. The polar histogram of ψ for all the pairs is shown in Figure 3.11(b), with the corresponding linear histogram in Figure 3.11(c). There is a bias in the distribution towards $\psi = 0$, confirmed by the Rayleigh and V tests (for circular uniformity) [205, 206] which reject the null hypothesis that the distribution is uniform at the 95 % level, i.e., there is a significant preference towards pair cells moving in the same direction. Example microscopy images of a pair moving in this way are shown in Figure 3.12.

Binning ψ according to the separation distance between two cells shows that this bias primarily occurs at small separations ($< 50\,\mu$m) as shown in Figure 3.13. The Rayleigh and V tests provide evidence to reject that each of the histograms in Figure 3.13 is uniform at the 95 % level. A measure of the skew is shown in the first moment, i.e., the arithmetic mean, $\overline{\psi}$ (as opposed to the circular mean). For a uniform distribution between 0 and π the arithmetic mean would be $\overline{\psi} = \pi/2$ or 90°. For the ψ distributions for $r < 20\,\mu$m, between 20–50 μm, between 50–100 μm and $r > 100\,\mu$m the arithmetic mean values are respectively, $\overline{\psi} = 73°$, 79°, 89° and 88°, indicating there is bias towards $\psi = 0$ at smaller separations.

Pearson's moment coefficient of skewness, $\gamma = E[(\psi - \overline{\psi})^3]/\sigma_\psi^3$, also provides a measure

of the asymmetry in the distributions. For a perfectly symmetrical distribution $\gamma = 0$, while for a distribution skewed towards lower values $\gamma > 0$ and for skew towards higher values $\gamma < 0$. For $r < 20\,\mu$m, between 20–50 μm, between 50–100 μm and $r > 100\,\mu$m the skewness values are respectively, $\gamma = 0.40, 0.24, 0.04$ and -0.02, showing reducing skewness towards $\psi = 0$. The Kolmogorov-Smirnov test provides no evidence to reject the null hypothesis that the distributions for $r < 20\,\mu$m and $20 < r < 50\,\mu$m are the same. Similarly for $50 < r < 100\,\mu$m and $r > 100\,\mu$m. However the test rejects the null hypothesis that the two smaller separation distributions are the same as the two larger separation distributions.

Calculating $\overline{\psi}$ with separations binned more frequently shows the length at which the movement is correlated, Figure 3.14. By performing a least-squares fit of the form $\overline{\psi} = 90[1 - e^{-(r+r_0)/m}]$, for parameters r_0 and m, we obtain the line $\overline{\psi} = 90[1 - e^{-(r+23.0)/25.9}]$ with an R^2 value of 0.6, shown in Figure 3.14. The characteristic length of the decay is therefore 26 μm, a typical cell diameter. We see that, on average, the cells farther than 70 μm show uncorrelated motion. This suggests that the previous estimate of 150 μm [87] could be an over-estimation of the interaction distance for pairs.

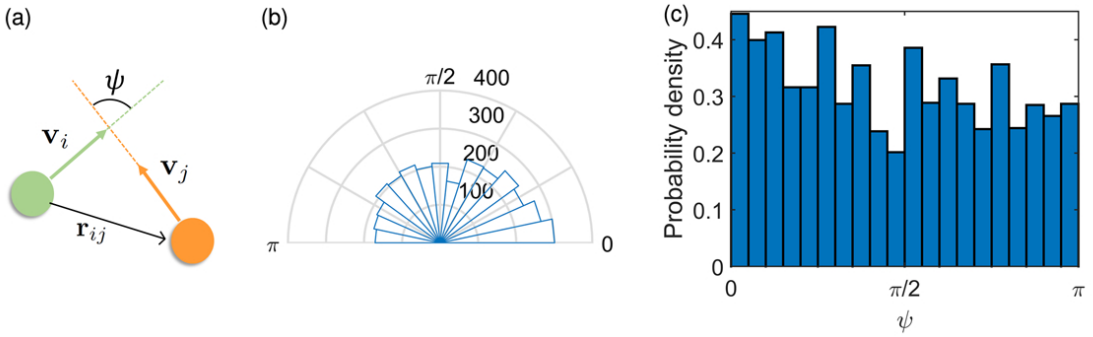

Figure 3.11: (a) Green and orange dots represent a pair of cells with their corresponding velocity vectors $\mathbf{v}_i$ and $\mathbf{v}_j$ together with their connection vector $\mathbf{r}_{ij}$. The angle between the velocity vectors is marked as ψ. (b) Polar histogram of ψ for all 50 pairs of cells with 3285 observations. (c) The corresponding linear histogram.

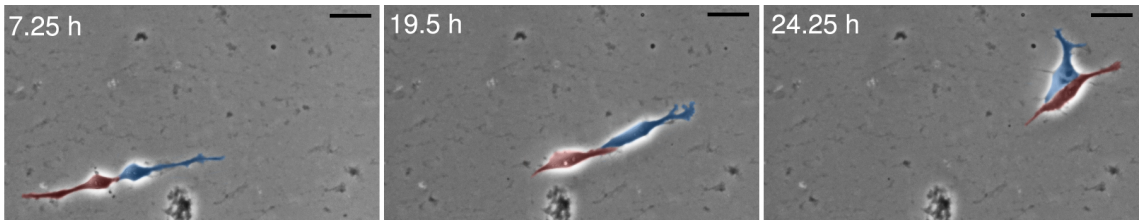

Figure 3.12: Example pair moving together in the same direction. The pair originates from single cell 17 (shown in the cell timeline in Chapter 2). The scale bar shows 20 μm.

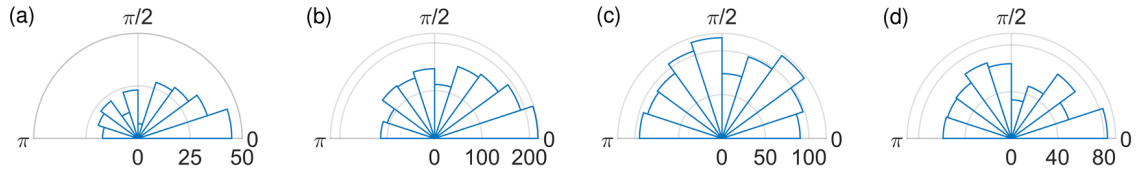

Figure 3.13: The angle between velocity vectors, ψ, for separations r (a) $< 20\,\mu$m, (b) 20–50 μm, (c) 50–100 μm, and (d) $> 100\,\mu$m, with 10 bins and 240, 1480, 974 and 591 measurements respectively.

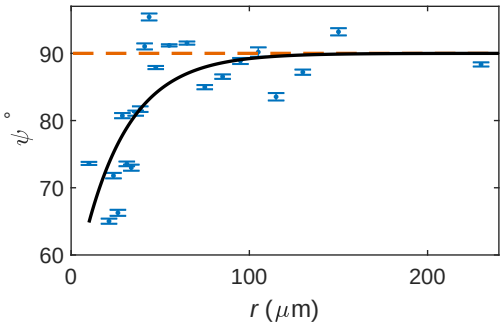

Figure 3.14: $\overline{\psi}$ binned by the separation distance, r, between two cells. Error bars show the standard error in the mean. The red dashed line shows 90 °, the expected value for uncorrelated motion. The least-squares fit (solid black line) is $\overline{\psi} = 90[1 - e^{-(r+r_0)/m}]$, with $r_0 = 23.0 \pm 29.0$ and $m = 25.9 \pm 15.6$ and an R^2 value of 0.6.

3.2.4 Movement and cell morphology

For the majority of pairs in our experiment, there is cell-cell contact via extended pseudopodia for a large portion (or all) of the motion. This morphological connection can be seen in the example cell pair in Figure 3.12 and in Figure 3.15. The latter pair is seen with connected pseudopodia at the start of the image recording, and the contact remains for the 18 hours before a cell division occurs.

Knowing this connection is present in the majority of the cell pairs provides biological background information to guide the interpretation of our movement analysis results. For example, the change in the direction of motion, θ, for each individual cell in a pair reveals a distinct preference towards a change in direction at each step with a global maxima at $\theta = \pi$. The local maxima at $\theta = 0$, indicative of persistent movement in the same direction as the previous step, is significantly less for pairs of cells than for single cells (where the peaks at 0 and π are comparable). This shows that changes of direction and the resulting back-and-forth motion is more common in cells in a pair than cells in isolation. It is likely the morphological pseudopodia connection between the pairs contributes towards this effect, creating a physical link between neighbouring cells and restricting cell movement. For the example cell shown in Figure 3.15, the two cells move back-and-forth, towards and away from each other repeatedly, while maintaining the pseudopodia connection.

The average separation between cells in a pair, $\overline{r}$, approaches an asymptotic value of around 70 μm. This distance is characteristic of the maximum length of the pseudopodia connection from qualitative observations of the microscopy images (the maximum length of the connection between the pair shown in Figure 3.15 is around 60 μm). As hESCs survive and proliferate more successfully with neighbours [188], the formation of the pseudopodia connection perhaps functions to restrict cell movement away from neighbours.

From our analysis of the angle between the directions of movement of each pair, ψ we see that the cells exhibit correlated motion when they are roughly a cell diameter apart or less. Again, this morphological connection could play a key role in facilitating these correlated movements between cell pairs.

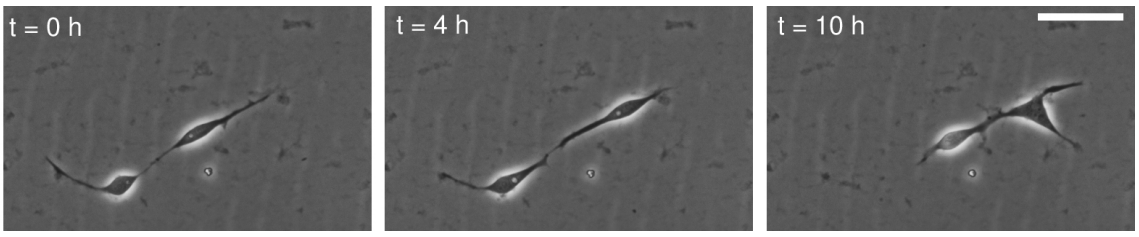

Figure 3.15: Microscopy images of a pair of cells with connected pseudopodia. The scale bar shows 50 μm.

3.3 Discussion and conclusions

Our findings suggest that the kinematics of hESC pairs are universal in terms of the individual cell and pair centroid speeds. Our results are within errors of previous measurements of the speeds of individual hESCs (at a low seeding density of 1500 cells/cm^2) of 25, 34 and 24 μm/h for cells that form a colony, join a colony and fail to form a colony, respectively [188]. Moreover, fibroblasts have significantly higher average speed of about 47 μm/h [188]. The hESC pairs have kinematic characteristics similar to those of individual cells. Thus, the kinematic diagnostics, if confirmed to be different for pluripotent cells and somatic cells, may lead to rapid, non-invasive diagnostics of the cell pluripotency and differentiation.

The cell pair as a whole (using the centroid position) undergoes a random walk with characteristic diffusivity at early times (< 12 h for unstained cells), with a similar motility to single cells ($D \approx 80\,\mu$m^2/h). The pairs show super-diffusive behaviour at later times, perhaps characteristic of a searching strategy, as suggested for single cells. These cells are more likely to negatively impact clonality, discussed further in Chapter 4. As for single cells, we see reduced motility for cells stained with the CellTrace™.

The pair separation over time increases exponentially before approaching an asymptote at 70 μm suggesting, on average, an optimal separation for pairs of cells. There is a preference for the cells to move in the same direction as each other on the plate at small separations ($< 50\,\mu$m). At these small separations it can be seen from the microscopy imaging that the cells are often physically connected by their pseudopodia. This coordinated movement could be due to an external stimulus, but it is likely the connection of the cell bodies facilitates this motion. Further experiments monitoring cell movement and morphology at increasing separations are needed to investigate whether this coordination is due to cell-cell contact alone, or whether any external biases or chemotaxis also contribute. At separations greater than $\approx 70\,\mu$m the motion of each cell in a pair appears uncorrelated, suggesting the previous estimate of 150 μm [87] could be an over-estimation of the interaction distance. Sometimes there is still a connection between the cell bodies at these distances, but the cells move in independent directions whilst maintaining the connection, and as an isotropic random walk when considered as a whole entity.

This additional information on low density plated cells will assist in the development of agent-based models, combining the motion of diffusive and super-diffusive cells with their biological states and cell-cell interactions. Establishing the diffusive migration of cells opens the possibility to use the well-developed, powerful mathematical theory of random walks for non-invasive prognostic modelling, with implications for large scale expansion and differentiation assays.

Chapter 4

Growing colonies of hPSCs

Stem cells form colonies through repeated divisions. Robust methods of producing cells under conditions that accelerate proliferation are particularly valuable as clinical applications of hPSCs require large numbers of cells ($\sim 10^{10}$). Additionally, for many of these clinical applications, clonal (genetically identical) cell populations are required [54, 55].

In this chapter we use experimental data to develop a stochastic exponential model for the growth of hPSC colonies, and use it to explore the impact of colony growth, leading to the merging of colonies with different progenitor cells, on clonality at different seeding densities. The model can be used to predict the critical time at which the clonality of the resulting colonies is compromised. The results we discuss here are also presented in [82].

4.1 Colony clonality

A typical *in-vitro* hPSC experiment involves the distribution of cells upon a growth material (the 'seeding' of cells onto a plate). The seeding density is then the initial number of cells placed on the growth material per unit area. Cells need to attach to the plate surface, which is covered by Matrigel or a similar substance (a gelatinous protein mixture which facilitates attachment), for viability and proliferation. However, some cells do not successfully attach and are lost. The hPSCs then form colonies by repeated mitosis in which two genetically identical daughter cells are produced from the division of the mother cell. The proliferation of cells in this way results in colonies of tightly packed cells in mono-layers along the growth material. The doubling time of stem cells varies and can be affected by various environmental and chemical factors, including the cell density [21, 31, 32].

An important measure of the self-renewal potential of stem cells is the *clonality*, the condition of being genetically identical. Generating homogeneous populations of clonal cells is of great importance [54, 55] as clonally derived stem cell lines maintain pluripotency and proliferative potential for prolonged periods [229]. Some applications require clonal homogeneous populations, e.g., drug discovery [14] and iPSCs for personalised medicine [230]. The selection of the best clones for further experimentation needs to be optimised to make clinical applications safe. If the seeding density is high, the migration of cells and the growth of closely-separated cell groups can cause aggregation of colonies; this is

undesirable when a homogeneous clonal population with identical genetic composition is required. The seeding density of cells has been shown to not only have an effect on the clonality of stem cells [87, 188], but also on their differentiation potential [231]. Moreover, culturing at an overly high density can cause DNA damage and culture adaptation, leading to increasing occurrence of chromosomal aberrations [21, 232, 233].

Single hPSCs are reported to have no effect on each other's movement if they are farther than 150 μm apart [87]. It is therefore recommended to keep a minimum distance of 150 μm between colony boundaries throughout growth to assure that the resulting individual colonies are each derived from a single founding hPSC. In Chapters 2 and 3 we considered the kinematics of single and pairs of cells, and identified occasional super-diffusive movements of cells which could lead to re-aggregation [88, 89]. An example of colony re-aggregation is shown in Figure 4.1.

As discussed in Chapter 1, the quantification of stem cell behaviours assists in the development of mathematical models which can then be used as non-invasive predictive tools. The development of colonies is inherently stochastic containing many properties which occur with some randomness, such as cell growth, cell division and cell death. We present a review of different modelling methods for colony proliferation in Chapter 1. In the next section we use experimental data to quantify the possibilities of re-aggregation due to the physical proximity of colonies at different seeding densities and consider the optimal seeding densities to form clonal structures.

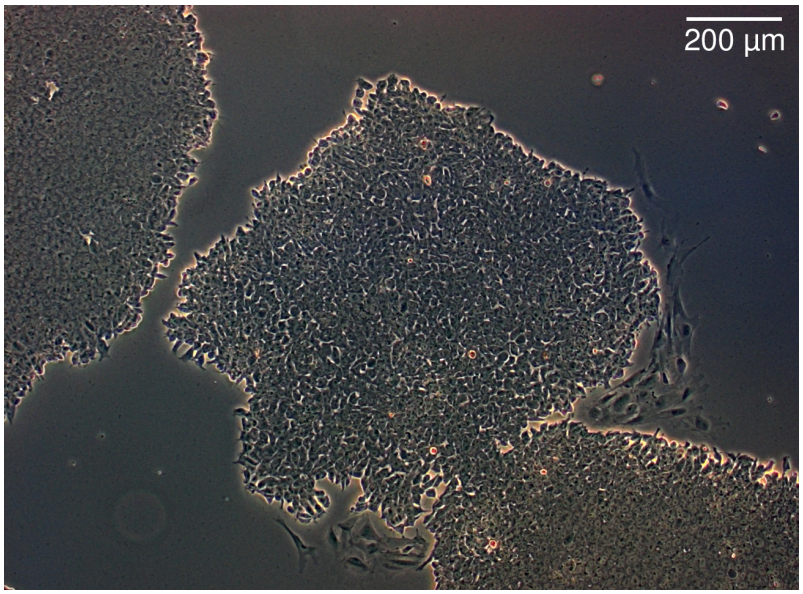

Figure 4.1: Microscopy image of colonies merging and thus losing the population clonality.

4.2 Colony growth

In this section we use experimental data of hESC colony populations to develop a stochastic growth model which can be used to calculate the critical time at which colonies re-aggregate and clonality is lost at different experimental seeding densities.

Carried out at the Institute of Biosciences at Newcastle University by Dr Irina Neganova. Data collection and image analysis by the author. Further experimental details are given in Appendix A.

Experiment 2a Single hESCs were plated at low density 1500 cells/cm^2 with microscopy images of the resulting 48 colonies taken at 72 h as shown in Figure 4.2. The population numbers of these colonies were extracted using IMARIS image analysis software through an automated procedure.

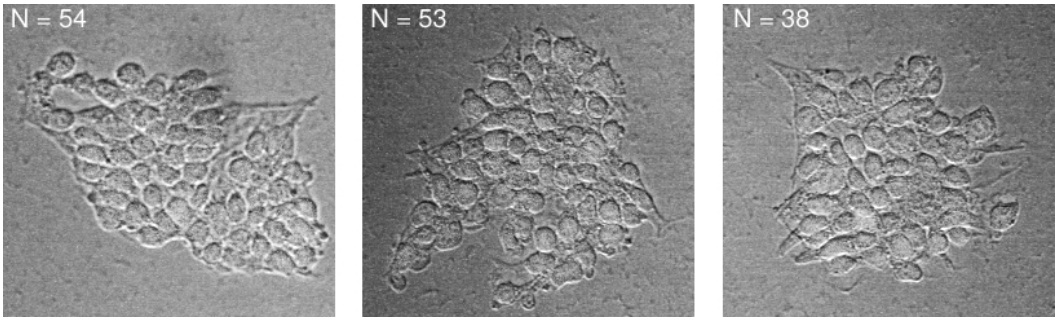

Figure 4.2: Microscopy images of colonies imaged at 72 h from Experiment 2a.

Experiment 2b Single hESCs were plated at varying seeding densities and the numbers of cells attached to the plate 24 h after cell seeding were counted. An example image of the seeded cells post-attachment is shown in Figure 4.3.

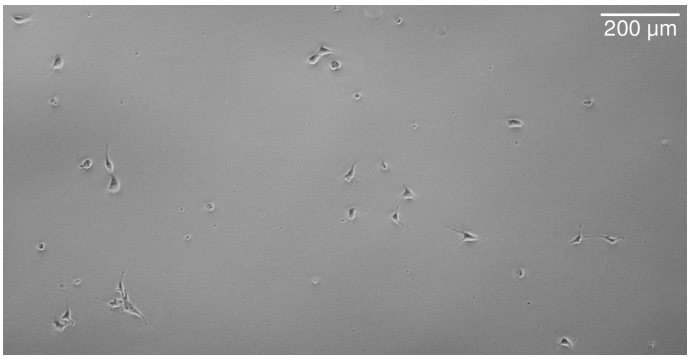

Figure 4.3: Attached cells seeded at 1200 cells/cm^2 from Experiment 2b.

4.2.1 Experimental populations

From Experiment 2a, the number of cells in each of the 48 colonies at 72 hours after cell attachment, $N(72)$, was recorded and is shown for each colony arranged in ascending order in Figure 4.4(a). The corresponding histogram of $N(72)$ is shown in Figure 4.4(b). The distribution is bimodal, illustrated by the kernel density estimation, with an outlier colony at $N(72) = 77$ cells. We remove this outlier colony for further analysis as it was most likely formed by several colony merges. Since the cell numbers were estimated using an automated process (through IMARIS software) several colony populations were also counted manually and the accuracy of the counting process verified (measurement error of around ± 2 cells).

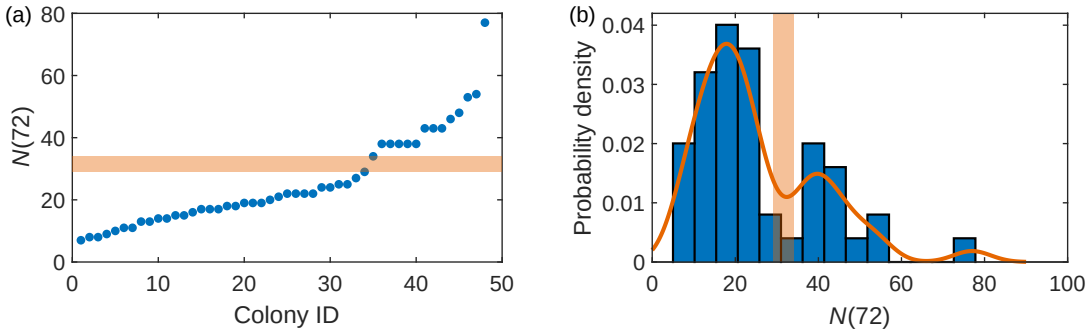

Figure 4.4: (a) $N(72)$ for each colony arranged in ascending order. The horizontal orange block shows the splitting of the data into two groups using K-means clustering. The split occurs between $29 \leq N(72) \leq 34$. (b) A histogram of $N(72)$ with kernel density estimate with a bandwidth of 4.5 (orange). The vertical orange block shows the splitting of the data into the two groups using K-means clustering.

We expect the number of cells at time t to evolve roughly as $N(t) = N_0 2^{t/t_\mathrm{d}}$, where $N_0 \equiv N(0)$ is the initial number of cells and t_d is the doubling time, i.e., the time it takes for the population to double, or equivalently $N(t) = N_0 e^{\gamma t}$, where γ is the growth rate. The bimodal nature of $N(72)$ implies we have two distinct groups of colonies, lead by differences in N_0 and/or differences in the growth rates between colonies. For a typical duration of the cell cycle, 16–18 hours, one expects 16–22 cells at 72 hours, corresponding to the first peak in the histogram in Figure 4.4(b). The second peak, at about 40 cells per colony, suggests that some groups of cells have merged to form larger colonies during the 72 hour period, or that the initial condition of the colony growth was in fact $N_0 = 2$. K-means clustering, a standard algorithm which partitions observations into clusters based upon minimising within cluster variance, splits $N(72)$ into the two ranges, $7 \leq N(72) \leq 29$ and $34 \leq N(72) \leq 77$.

To ascertain the initial conditions that underlie the colony growth, we turn to Experiment 2b examining the cells after 24 hours; a typical image is shown in

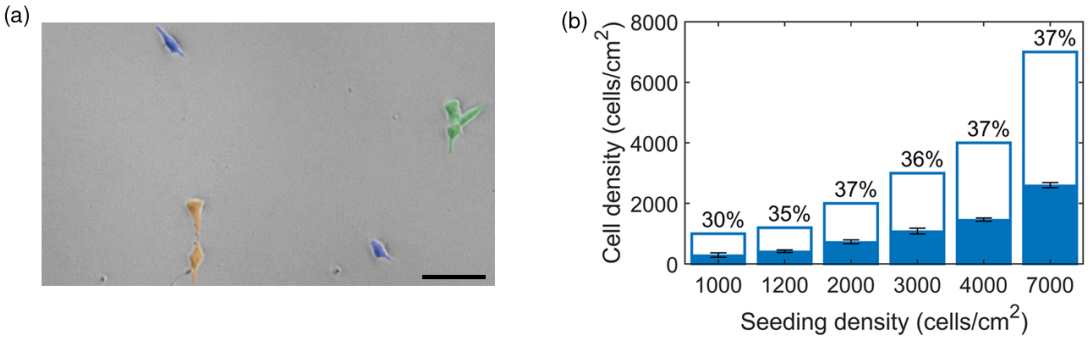

Figure 4.5: (a) Example microscopy image of cells seeded at density $n_0 = 1200\,\text{cells/cm}^2$. We see that some cells have no neighbours within $150\,\mu$m (highlighted in blue), but some are in pairs (orange) and triples (green). The scale bar represents $100\,\mu$m. (b) Bar chart showing the initial cell seeding density (cells/cm^2) at day zero (n_0, unfilled bar height), and the mean cell density of the remaining cells attached to the plate at day one (η_0, filled bars). Each mean is based on three independent measurements and the error bars represent standard deviations.

Figure 4.5(a). There are several characteristic features of the cell distribution revealed in this experiment. Firstly, the random initial positioning of the seeded cells means that some cells are initially isolated with no cells within the interaction distance (estimated as $150\,\mu$m). Other cells lie within the interaction distance of each other, forming groups of varying sizes. In Figure 4.5(a) we colour-code the cells according to whether they are isolated, or are effectively in a pair or in a triple, to illustrate how N_0 can vary at low seeding densities.

Secondly, only a fraction of the originally seeded cells are attached to the plate at this time. Cells need to be attached to the plate for viability and proliferation – cells which do not attach are lost. We find for a range of seeding densities, $1000 \leq n_0 \leq 7000\,\text{cells/cm}^2$, that on average $35.19\% \pm 4.23\,(0.99)$ (the mean $\pm$ one standard deviation (standard error)) of initially seeded cells were attached 24 hours after plating. The proportion of attached cells at different seeding densities is shown in Figure 4.5(b). This rate of attachment is usual for stem cells in similar experiments. In Section 4.2.2 we discuss $N(t)$, the number of cells present in a colony over time, beginning with N_0, independent of original cell seeding densities. In Section 4.2.3 we discuss the effects of the cell plating densities, n_0, where we assume that the actual density of cells present is $0.35n_0$ to account for the loss at the attachment stage. Note this relationship is easily adjustable for different attachment rates.

Throughout the next sections we define time $t = 0$ to be the time that seeded cells have attached to the plate and their proliferation starts. Before this time there will have been a process of attachment to the plate during which some cells were lost, as shown

in Figure 4.5(b), plus a delay in the growth from the lag-phase experienced by newly plated *in-vitro* hESCs as they adjust to the environment [234]. This is consistent with our experimental data which considers 72 hours after cell attachment.

4.2.2 Stochastic growth model

The simplest deterministic model for the number of cells in a colony at time t, $N(t)$, assumes a constant cell division time $1/\gamma$ and simultaneous division of all the cells, leading to

$$\frac{dN(t)}{dt} = \gamma N, \tag{4.1}$$

which has the solution

$$N(t) = N_0 e^{\gamma t}, \tag{4.2}$$

where $N_0 \equiv N(0)$ is the initial number of cells at $t = 0$. However, the cell cycle duration is variable due to various factors, such as inhomogeneities in the nutrient distribution within the growth medium and the inherent variation in the cell cycle duration between different clones. Such effects can be allowed for by considering a Gaussian random growth rate γ, with a mean value μ and standard deviation σ:

$$N(t) = N_0 e^{\gamma t}, \;\; \gamma \sim \text{Norm}(\mu, \sigma^2). \tag{4.3}$$

Different colonies thus grow at different rates sampled from the Gaussian probability distribution. Using the properties of normal distributions, we can then derive that the number of cells $N(t)$ follows a lognormal distribution. We can rewrite Eq (4.3) as $N(t) = e^{\log(N_0)+\gamma t}$ and since γ follows a normal distribution, it follows that $X \equiv \gamma t + \log(N_0) \sim N(t\mu + \log(N_0), t^2\sigma^2)$ through linear transformation properties. Therefore the exponential of the variable follows a lognormal distribution with the same parameters (by the definition of a lognormal random variable), $e^X \equiv N(t) \sim \text{Lognormal}(\mu_0, \sigma_0^2)$ where $\mu_0 = t\mu + \log(N_0)$ and $\sigma_0^2 = t^2\sigma^2$. The parameters μ_0 and σ_0 can be estimated from the experimental data.

Assuming that each colony begins as a single cell, i.e., $N_0 = 1$, maximum likelihood parameter estimation gives $N(72) \sim \text{Lognormal}(3.08, 0.56^2)$, shown in Figure 4.6(a). The model based on the experimental data therefore becomes Eq (4.3) with $\mu = 0.0428\,\text{h}^{-1}$ and $\sigma = 0.0077\,\text{h}^{-1}$, corresponding to a doubling time of approximately 16 hours. The overall shape of the distribution in Figure 4.6(a) is roughly captured but the bimodal nature is not. It is worth noting that when simulated stochastically this single population model can result in a bimodal distribution for $N(t)$ given the probabilistic nature of the stochastic simulations, see Figure 4.6(b), but this is a rare occurrence happening around once in every 100 simulations for 50 colony populations. Note that simulating from Eq (4.3) will always recover exactly the theoretical lognormal distribution when ran for large enough

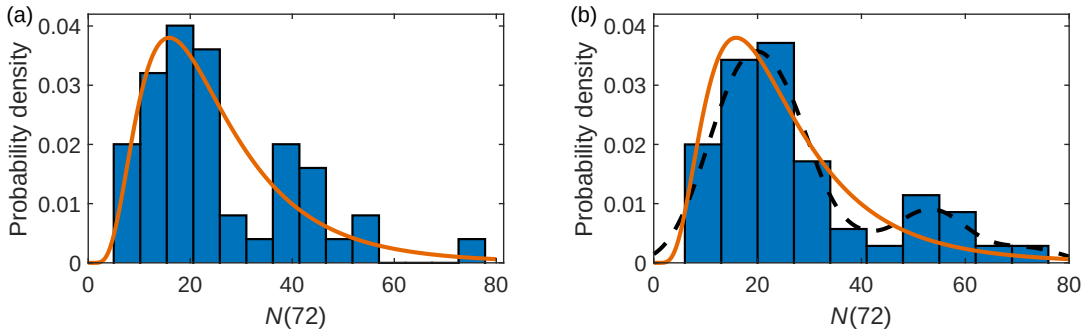

Figure 4.6: (a) Experimental $N(72)$ histogram with the fitting obtained from Eq (4.3), $N(72) \sim \text{Lognormal}(3.08, 0.56^2)$. (b) An example simulation from Eq (4.3) for 50 colonies at 72 hours producing a bimodal distribution, with kernel density estimation (black dashed) and the theoretical lognormal fitting (orange).

numbers of colonies. This bimodal effect is due to the stochasticity at smaller sample sizes (in this case 50 colonies).

We suggest that a bimodal distribution of the colony sizes can be a consequence of a difference in the cell proliferation rates in cell groups of different sizes that may arise from their interactions. It can be expected that colonies starting from larger groups grow faster, not only due to the initial conditions but the preference of cells to growth in close proximity to neighbours [52, 188, 235]. To capture the bimodal nature of the colony size distribution, we consider two populations, A and B, each with a different initial condition

$$N_0 = \begin{cases} 1, & \text{group A}, \\ 2, & \text{group B}, \end{cases} \tag{4.4}$$

where the probability of a colony belonging to groups A and B are α and β, respectively. Each population then follows Eq (4.3) with its corresponding initial condition,

$$\begin{cases} N_{\text{A}}(t) = e^{\gamma_{\text{A}} t}, \;\; \gamma_{\text{A}} \sim \text{Norm}(\mu_{\text{A}}, {\sigma_{\text{A}}}^2), & \text{with probability } \alpha, \\ N_{\text{B}}(t) = 2e^{\gamma_{\text{B}} t}, \;\; \gamma_{\text{B}} \sim \text{Norm}(\mu_{\text{B}}, {\sigma_{\text{B}}}^2), & \text{with probability } \beta, \end{cases} \tag{4.5}$$

and each of N_{A} and N_{B} has a lognormal probability distribution. Thus we consider separately colonies that originate from a single cell and those from cell pairs. We consider the possible role of cell triples and larger progenitor groups in Section 4.2.3.

A lognormal mixture fit to the data, shown in Figure 4.7(a), gives $N_{\text{A}}(72) \sim \text{Lognormal}(2.84, 0.41^2)$ and $N_{\text{B}}(72) \sim \text{Lognormal}(3.76, 0.13^2)$, with the mixture probabilities $\alpha = 0.77$ and $\beta = 0.23$. Therefore, we have $\mu_{\text{A}} = 0.0394\,\text{h}^{-1}$, $\sigma_{\text{A}} = 0.0057\,\text{h}^{-1}$, $\mu_{\text{B}} = 0.0426\,\text{h}^{-1}$, and $\sigma_{\text{B}} = 0.0018\,\text{h}^{-1}$. A comparison between these growth parameter distributions is shown in Figure 4.7(b). These values correspond to a

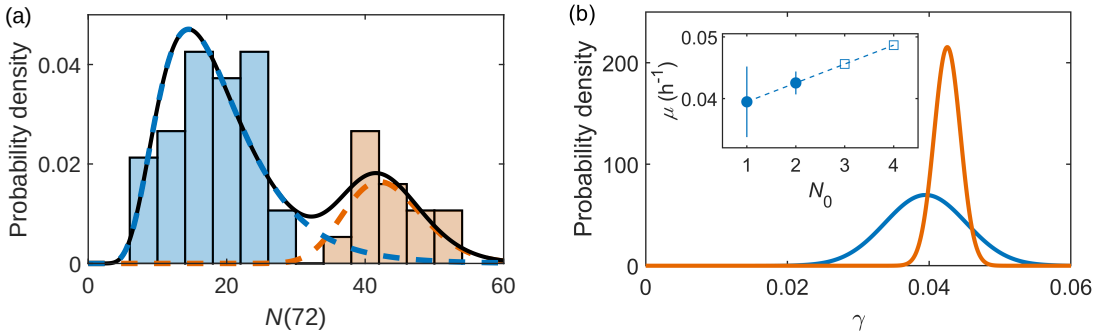

Figure 4.7: (a) $N(72)$ with a lognormal mixture model fitting corresponding to the two populations. The population fittings, $N_{\mathrm{A}}(72) \sim \mathrm{Lognormal}(2.84, 0.41^2)$ with mixture probability 0.77 and $N_{\mathrm{B}}(72) \sim \mathrm{Lognormal}(3.76, 0.13^2)$ with mixture probability 0.23 are shown in dashed blue and orange respectively, with the overall mixture distribution in black. (b) The two γ distributions for group A (blue) and group B (orange). Note that the distributions are not scaled to represent the group probabilities α and β. The inset shows the parameters for γ, μ with $\pm\sigma$ error bars for the initial conditions $N_0 = 1$ and $N_0 = 2$, corresponding to group A and group B, respectively. The dashed line shows the extrapolation of the trend to higher values of N_0.

doubling time of 17.5 h for the single founder cell population (group A) and 16.3 h for the pairs of cells population (group B), in line with other experimental doubling times for hESCs [87, 88, 188]. Note that here we have assumed the growth rates for the two populations are different. Our fitting suggests colonies starting with two cells grow faster than their single cell counterparts, consistent with the idea that cells growing in larger colonies proliferate more effectively than isolated cells [52, 188, 235]. The fitting under the assumption of identical growth rates, $\gamma_{\mathrm{A}} = \gamma_{\mathrm{B}}$, is shown in Figure 4.8. This fitting does not capture the experimental data and so we conclude it is more appropriate to continue assuming the two populations have different growth rates, $\gamma_{\mathrm{A}} \neq \gamma_{\mathrm{B}}$.

We have demonstrated how accurately the two-population model captures the experimental data at 72 hours. Now we proceed to develop it into a prognostic model for the colony size at later times. The evolution of the colony size, $N(t)$, according to this two-population model is shown in Figure 4.9. Because of the random scatter in the colony growth parameters the admissible range of colony size $N(t)$ increases as $\sqrt{t}$, and sooner or later, the size of the two colony types overlap. At early times, the sizes of the two colony types are distinct, where those beginning from two cells are larger than those from one founder cell, but as time progresses the stochasticity in the growth rates causes an overlap in the two populations. This overlap becomes more prominent for larger numbers of simulations as this increases the incidence of extreme growth rate values. Histograms of $N(20)$ and $N(72)$ in Figures 4.9(e) and 4.9(f) illustrate how at early times the two colony types are distinguishable but over time the distributions spread and

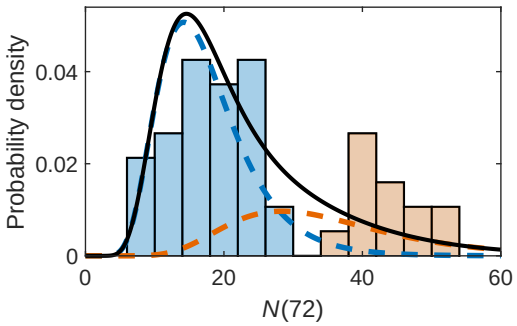

Figure 4.8: The mixture fitting assuming the growth rates for the two populations are the same. The population fittings, $N_{\mathrm{A}} \sim \mathrm{Lognormal}(2.79, 0.61^2)$ with mixture probability 0.72 and $N_{\mathrm{B}}(72) \sim \mathrm{Lognormal}(3.49, 0.61^2)$ with mixture probability 0.28 are shown in dashed blue and orange respectively, and the overall mixture distribution in black.

merge to make single-clone colonies indistinguishable from heterogeneous colonies.

The time at which N_{A} first becomes equal to N_{B}, t_*, is the critical time after which it is not possible to distinguish which colonies originated from a single progenitor based on the colony size. The mean of this time is shown for increasing numbers of colonies in Figure 4.10. As we increase the number of colonies, N_{cols}, we see more of the extreme values occurring with low probability, causing the blurring of the two populations to begin at an earlier time. The relationship is a power law, with the best fit $t_* = a{N_{\mathrm{cols}}}^{-b}$ with $a = 77.9 \pm 4.7\,\mathrm{h}$ and $b = 0.12 \pm 0.01$ with an R^2 coefficient of 0.98. This allows us to estimate, for a given plating cell density, the time up to which colonies originating from a single founder cell are identifiable based on the current number of cells in the colony.

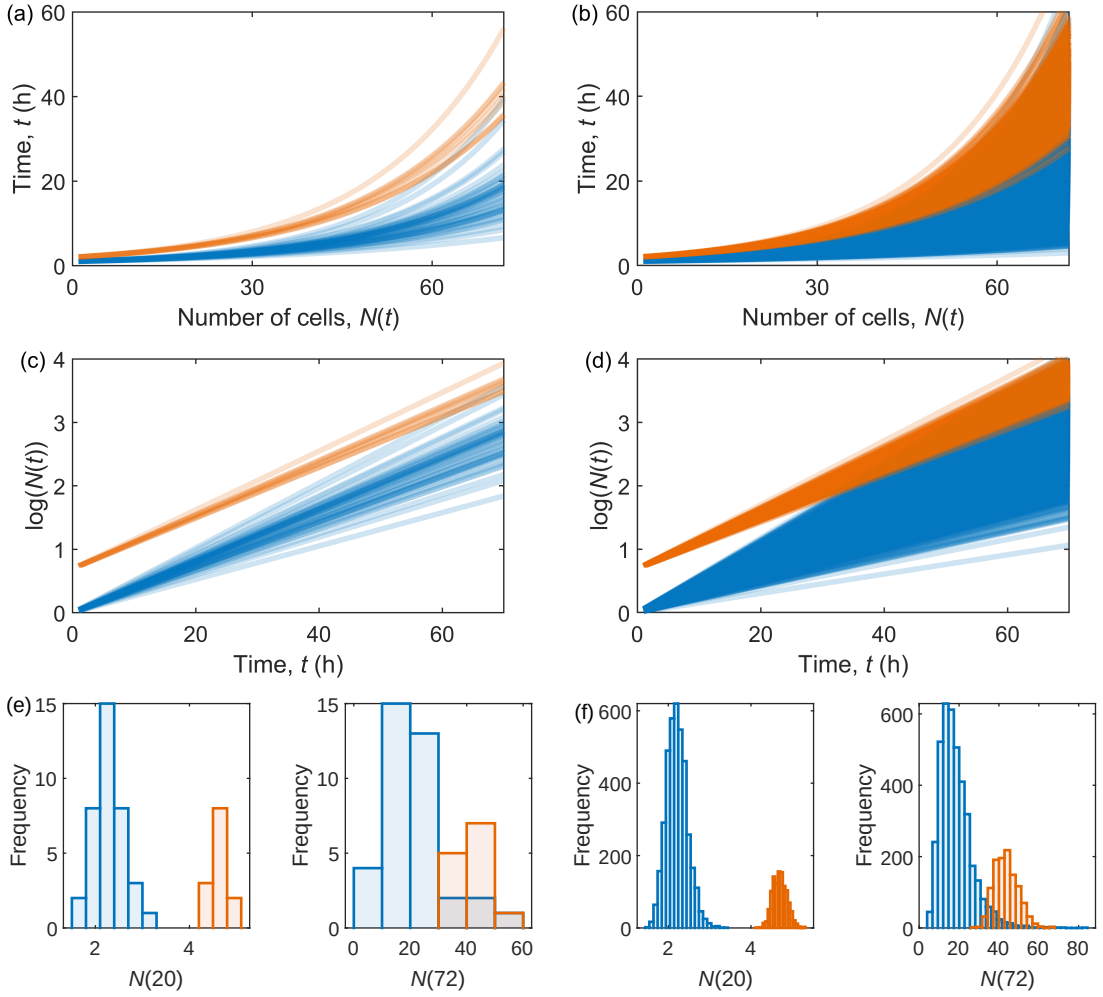

Figure 4.9: Evolution of $N(t)$ from the multi-population model Eq (4.5) for (a) 50 and (b) 5000 colonies, with blue N_A ($N_0 = 1$) and orange N_B ($N_0 = 2$) colonies. (c) and (d) show the corresponding plots of $\log(N(t))$ over time. Histograms of $N(20)$ and $N(72)$ for (e) 50 and (f) 5000 colonies. Note that the distributions are unscaled to highlight the difference in the frequencies of each group, resulting from the difference in the population probabilities $\alpha = 0.77$ (blue population) and $\beta = 0.23$ (orange population).

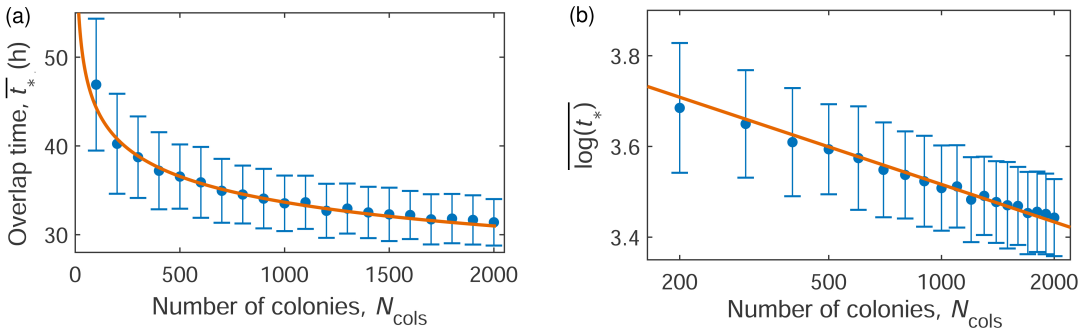

Figure 4.10: (a) The mean critical time, $\overline{t_*}$ at which colonies originating from a single cell are no longer identifiable based on colony size, $N(t)$. Each data point is the mean of 500 simulations. The error bars represent the standard deviations in the means. The data is well captured by a power law relation $\overline{t_*} = a{N_{\text{cols}}}^{-b}$ (orange line, $R^2 = 0.98$) with (b) linear least-squares fitting on a log-log plot (inset) giving $a = 77.9 \pm 4.7\,\text{h}$ and $b = 0.12 \pm 0.01$.

4.2.3 Cell clustering and seeding density

Typical low seeding densities for hESCs, intended to grow colonies from single founder cells, commonly range from 500 to 3000 cells/cm^2. Across a range of seeding densities, we find that the average proportion of cells attached to the substrate after 24 hours is $35 \pm 4\%$, where the range represents one standard deviation within the sample, and the accuracy of the mean value (the standard error) is $\pm 1.0\%$, presented in Figure 4.5(b). For example, an initial seeding density of $n_0 = 1500$ cells/cm^2 results in around 500 cells/cm^2 continuing past day one of the experiment. Throughout this section we will present the initial seeding densities n_0 and work on the assumption that 35% of these cells are successfully attached and survive, $\eta_0 = 0.35 n_0$.

In this section we find the initial conditions corresponding to cell plating at different cell seeding densities and use this to inform the model for colony growth. The seeding of cells randomly across a growth area, A, can be simulated as a homogeneous Poisson point process in which the number of point counts is sampled from a Poisson distribution with the mean λA, where λ is the point intensity. The points are then independently and uniformly scattered across the region. For example, if we consider an average initial seeding density of $n_0 = 1500\,\text{cells/cm}^2$ (and $\eta_0 = 0.35 n_0$), we can simulate the estimated number of cells (point counts) post-attachment in an area A as $N_0 \sim \text{Po}(\eta_0 A)$. The N_0 cells can then be located randomly according to uniform distributions in x and y.

Once cells have been scattered, we can then consider the distances between cells and their nearest neighbours with the aim to estimate the fraction of isolated cells, their pairs (defined as two cells separated by less than $150\,\mu$m) and triples etc. The probability density function of the distance, r, to the k^{th} nearest neighbour is known from the theory of Poisson point processes as $d_k(r) = 2(\lambda\pi r^2)^k e^{-\lambda\pi r^2}/r(k-1)!$ [236]. This reduces to

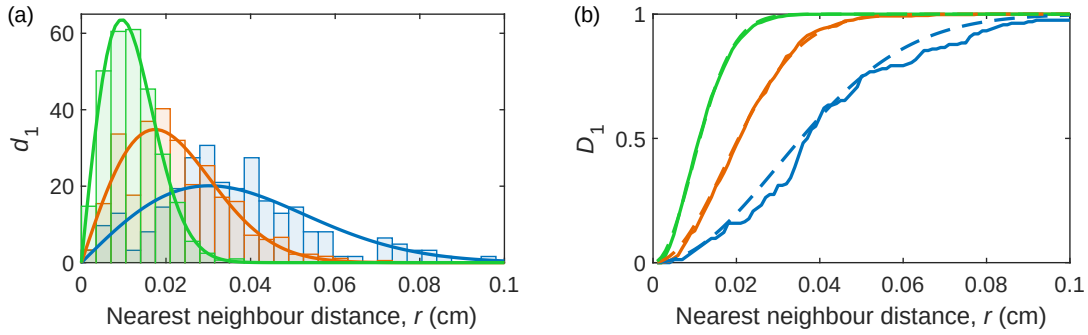

Figure 4.11: (a) The probability distributions for the nearest neighbour distance in a Poisson point process, given by $d_1(r) = 2\lambda\pi r e^{-\lambda\pi r^2}$ for $\lambda = \eta_0 = 175$ (blue, $n_0 = 500\,\text{cells/cm}^2$), $\lambda = \eta_0 = 525$ (orange, $n_0 = 1500\,\text{cells/cm}^2$) and $\lambda = \eta_0 = 1750$ (green, $n_0 = 5000\,\text{cells/cm}^2$). The histograms show the simulated distributions for seeding cells according to the process, each is for one run of the seeding simulation. (b) The theoretical proportion of cells with a nearest neighbour $< r$ cm away, $D_1(r) = 1 - e^{-\lambda\pi r^2}$, for $\lambda = \eta_0 = 175$ (blue dashed, $n_0 = 500\,\text{cells/cm}^2$), $\lambda = \eta_0 = 525$ (orange dashed, $n_0 = 1500\,\text{cells/cm}^2$) and $\lambda = \eta_0 = 1750$ (green dashed, $n_0 = 5000\,\text{cells/cm}^2$). The corresponding solid lines show the values of D_1 from a single run of the seeding simulation.

$d_1(r) = 2\pi\lambda r^2 e^{-\lambda\pi r^2}$ for the first nearest neighbour. The theoretical distributions along with histograms from simulated data for $d_1(r)$ are shown in Figure 4.11(a) for initial seeding densities of $n_0 = 500$, 1500 and $5000\,\text{cells/cm}^2$ corresponding to $\lambda = \eta_0 = 0.35 n_0$. These distributions allow us to calculate the proportion of seeded cells with the nearest neighbour at a given distance. The nearest neighbour cumulative distribution function for the proportion of cells with a nearest neighbour at a distance $< r$ for a 2D homogeneous Poisson process is given by $D_1(r) = 1 - e^{-\lambda\pi r^2}$ [236]. This theoretical proportion of cells with a first nearest neighbour less than r away, D_1, for initial seeding densities $n_0 = 500$, 1500 and $5000\,\text{cells/cm}^2$ is shown in Figure 4.11(b) along with data from a simulation at each seeding density. For the initial seeding density of $n_0 = 1500\,\text{cells/cm}^2$ the nearest neighbour distance between cells will be less than $150\,\mu$m in around 30% of cases, similar to the experimental estimate of 23%. We have neglected the movement of cells as, based on observed migration speeds of approximately $18\,\mu$m/h [88], the time required to traverse the critical interaction distance of $150\,\mu$m is around 9 h, a large portion of the cell cycle.

To consider the groupings of seeded cells we use a density based clustering algorithm. Cells less than $150\,\mu$m apart are considered as being part of the same cluster, and any neighbouring cell less than $150\,\mu$m away from any other cell in the cluster is also considered to be part of the cluster. This allows for clusters of elongated shapes. This definition of a cluster of cells is non-trivial and has implications for the interactions of cells, but from the experimental images we see examples of clusters in elongated shapes as well as the more common regular shapes. Examples of different cluster formations for triples are shown in

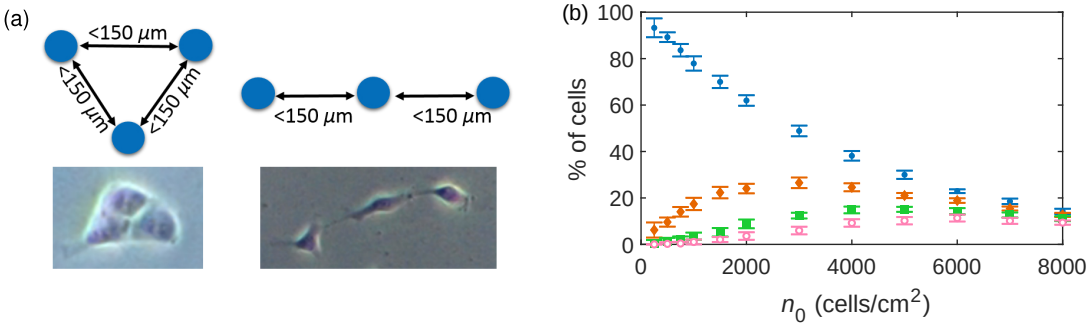

Figure 4.12: (a) Example clustering configurations for cell triples with experimental formations. The distance between cells, d, must be less than 150 μm. (b) The percentage of cells in a cluster of size n (for an interaction radius of 150 μm) for $n = 1$ (single cells, blue circles), $n = 2$ (pairs, orange diamonds), $n = 3$ (triplets, green square) and $n = 4$ (quadruples, pink open circles) with standard deviation error bars.

Figure 4.12(a). The proportion of cells in a cluster of a size n at different seeding densities is shown in Figure 4.12(b). At low initial seeding densities, e.g., $n_0 = 500\,\text{cells/cm}^2$ the majority of cells have no close neighbours. As the seeding density increases, the proportion of pairs of cells increases. The proportion of each cluster size first rises with n_0 before reaching a maximum and then tends to zero as more possible cluster sizes become available. The distributions shown in Figure 4.12(b) provide the initial conditions corresponding to cell seeding at different densities.

Now the initial conditions of cell seeding are known, the growth of colonies from these cells can be considered. Cells are seeded at density n_0 according to a Poisson point process as described above, and then the division of the cells and growth into colonies can be described by the two-population model, Eq (4.5). The area coverage of the plate can be estimated from the number of cells we expect to be present. The average area of a cell, A_{cell}, is approximately 250 μm^2 corresponding to a cell diameter of 18 μm [104]. The percentage of area covered by cells evolves as shown in Figure 4.13(a). Taking this value, the proportion of area coverage is the area covered by the cells, $N(t)A_{\text{cell}}$, divided by the growth area of the plate and $A_{\text{plate}} = N_{\text{seeded}}/n_0$. We therefore expect the percentage area coverage over time to tend to an exponential relationship due to the growth of $N(t)$, scaled by a factor equal to $n_0 A_{\text{cell}}/N_{\text{seeded}}$, as we see in Figure 4.13(a). The time taken for the growth area to be 100% covered, t_{A100}, for varying initial seeding densities, is shown in Figure 4.13(c).

We can now simulate the growth of colonies spatially using the initial conditions as described above and the colony growth model from Eq (4.5). This allows us to estimate the crucial time at which the colonies begin to merge. The cells are seeded at density n_0, with η_0 cells attached, and are then sorted into clusters based on their spatial distances away

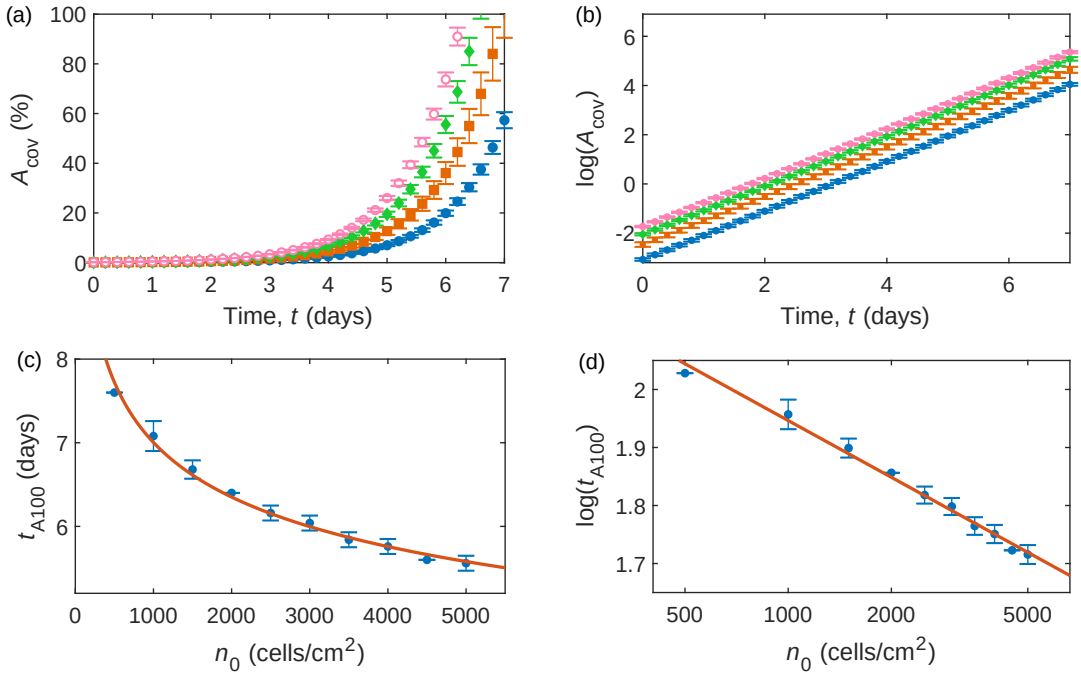

Figure 4.13: (a) The percentage area coverage of cells at different time points for different cell seeding densities (blue 500 cells/cm^2, orange 1000 cells/cm^2, green 1500 cells/cm^2 and pink 2000 cells/cm^2) according to the two-population model. (b) Corresponding data on a log-linear scale. (c) The time the growth area is 100% covered for varying cell seeding densities, with line of best fit $t_{A100} = 20n_0^{-0.15}$. (d) The data on a log-log scale, with line of best fit $\log(t_{A100}) = (-0.149 \pm 0.005)\log(n_0) + (2.966 \pm 0.04)$.

from each other. Each cluster grows according to the two-population model, approximated as a circle with centre at the geometric centre of the cluster and radius based on the number of cells present, $N(t)$. The growth rate for triples and larger clusters of cells is assumed to be the same as that for pairs (although this assumption may have to be revised when more detailed experimental data is available). The time at which any colony begins to merge with its neighbour is the critical time that clonality is lost, τ, illustrated in Figure 4.14(a). The time the first colony merge occurs at varying seeding densities is shown in Figure 4.14(b), with least squares fitting $\tau = (-0.007 \pm 0.0001)n_0 + (102 \pm 3)$ with $R^2 = 0.99$, τ in hours and n_0 in cells/cm^2. We can then estimate the time the first colony merge will occur from the equation

$$\frac{\tau}{1\,\text{h}} \approx 100 - \frac{n_0}{140\,\text{cm}^{-2}}, \tag{4.6}$$

where n_0 is the initial seeding density of cells before attachment in cells/cm^2 and τ is produced in hours. Experimental values were extracted for τ from Experiment 2 and the model captures these values within errors for the seeding densities 3000, 4000 and

Seeding density, n_0 (cells/cm^2)	Attached cells, η_0 (cells/cm^2)	% of single cells	Time to first colony merge, τ (h)
1000	350 ± 42	78 ± 3	94 ± 3
1500	525 ± 63	70 ± 3	91 ± 3
3000	1050 ± 127	48 ± 1	80 ± 5
5000	1750 ± 212	30 ± 1	67 ± 6

Table 4.1: The expected number of attached cells, single cells at seeding and the time to the first colony merge for varying seeding densities.

7000 cells/cm^2. These results are summarised in Table 4.1 for convenience. The results are also shown for extrapolated growth rates in Figure 4.14(b), under the assumption growth rates continue to increase with cluster size as in Figure 4.7(b). The least squares fitting is $\tau = (-0.01 \pm 0.001)n_0 + (97 \pm 4)$ with $R^2 = 0.99$, τ in hours and n_0 in cells/cm^2. The increasing growth rates cause an earlier first merging time, particularly at the higher seeding densities where larger clusters are more likely.

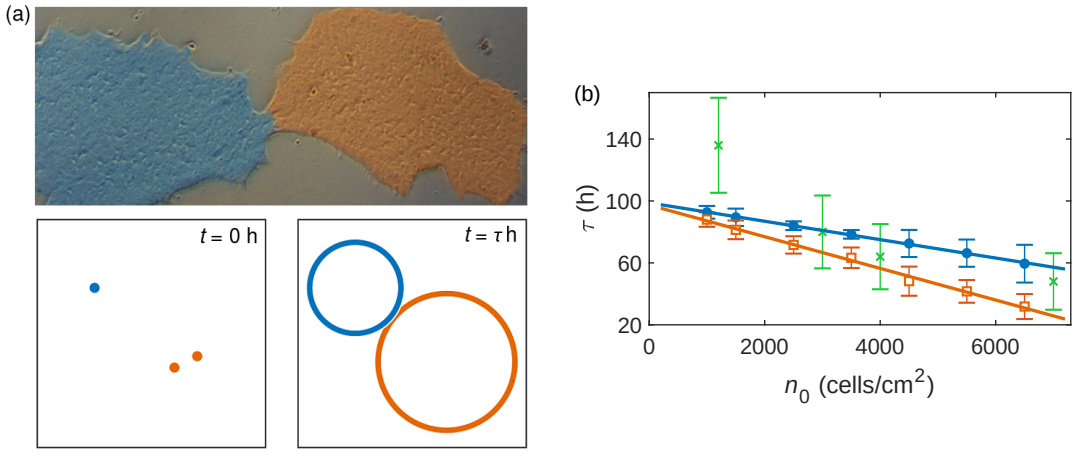

Figure 4.14: (a) An example of two colonies merging together at day 5 from experimental images. The below diagram illustrates initially seeded cells and the colonies at the first merge time τ from a simulation of the cell seeding model. The orange cells are classed as a pair and grow accordingly. (b) The mean time (averaged over 20 simulations) that the first colony merge occurs for cells seeded at different densities growing according to the two-population model, Eq (4.5). The time assuming the growth rates for clusters of three or above are the same as the growth rate for pairs is shown as blue circles with the line of best fit, $\tau = (-0.007 \pm 0.001)n_0 + (102 \pm 3)$ with $R^2 = 0.99$ in blue. The mean time using extrapolated growth rates for larger clusters is shown as orange squares with line of best fit $\tau = (-0.01 \pm 0.001)n_0 + (97 \pm 4)$ with $R^2 = 0.99$ in orange. To test the prediction, experimental values of τ were extracted from Experiment 2b and are shown as green crosses. The error bars have been calculated through error propagation based on an error of ±0.5 days for each of the measured values due to the image time resolution.

4.3 Discussion and conclusions

Colony growth originating from single or pairs of cells is well modelled using a two-population stochastic exponential model where the growth rate is sampled from a Normal distribution. Experimental results show that colonies that start from a single founder cell grow at a rate different from those originating from a cell pair, with a relative difference of $8 \pm 1.8\%$. The colonies originating from pairs of cells have a higher mean growth rate (although it is within standard deviation errors of the growth rate for single cell colonies), and the standard deviation of the growth rate for pairs is much lower, as seen in Figure 4.7(b).

Stem cells do not show any contact inhibition of proliferation which would slow colony growth, however, it is worth noting that there could be a trend in growth rates with respect to the 'tightness' of the cluster packing. It is likely that a triplet of cells in contact grows faster than a triplet where each cell is 150 μm apart, based on the evidence that stem cells proliferate more effectively in close proximity with each other [52, 188, 235]. It is likely that the growth rates are actually time dependent, so although we see a difference in the growth rates from single and paired founding state up to 72 h, we expect this difference to reduce and eventually become negligible over time. Similarly, we expect that the effect of the difference in growth rates between triples and other larger groups of founding cells is also diminished as the cluster size increases. It is for this reason we can set the growth rate for clusters of triples and larger groups to be the same as the growth rate for pairs. Allowing the growth rates to continue increasing with cluster size (extrapolated from the data) results in less time to the first colony merge, especially at the higher seeding densities where the probabilities of larger clusters is increased, Figure 4.14(b). Further experimental data (colony population sizes with time) would be needed to clarify the trends in growth rates for both increasing cluster size and cell separation distance within the cluster.

This model can be used to predict colony sizes at different time scales. We thus suggest that single clone colonies (that originate from a single founder cell) can be selected as the smaller of the growing colonies, but only within a certain period, t_*, before colony sizes from a range of founder cells can become similar to each other. This time gives an indication of when colonies should be observed to identify single founder cell colonies based on colony size. This time, t_*, ranges from 30–50 h depending on the seeding density.

Experimental results show that on average 35% of initially seeded cells are attached to the plate 24 hours after cell seeding, Figure 4.5(b). We take this into account when considering the initial conditions of cell seeding. It is thought that cells affect each other's movement if they are less than 150 μm apart [87]. The nearest neighbour distributions for cells seeding according to a Poisson process, shown in Figure 4.11, indicates that for a low seeding density of $n_0 = 1500\,\text{cells/cm}^2$ (i.e., around 500 attached cells) we would expect to

see around 30% of cells with a nearest neighbour closer than the critical value $r = 150\,\mu$m. However, our results from Chapters 2 and 3 suggest that we would not expect all these pairs to migrate towards each other and form colonies [88, 89]. This critical separation distance may be affected by the cluster size of neighbouring colonies. Since the cellular communication is by inter-cellular signalling, it is likely reduced as neighbouring colony sizes increase. At higher seeding densities this individual cell migration is more likely to have an effect on the cell clustering at early times. However, as larger groups of cells in close contact form colonies, individual migration reduces to a negligible amount, meaning it is only a factor at early times. For large colonies there could be collective migration effects to consider. Microscopy images of large colonies moving over a growth area would be needed to quantify this movement, which could then be introduced into the model.

The initial clustering of cells is an important and non-trivial consideration with implications on cell communications. We define a cluster to be a group of cells where each cell is $150\,\mu$m or closer to another cell in the cluster. This allows clusters of different structures, as illustrated in Figure 4.12(a). This random seeding process gives the proportion of cell clusters of different sizes presented in Figure 4.12(b). At low seeding densities up to $\approx 3000\,\text{cells/cm}^2$, i.e., ≈ 1000 attached cells in a typical experiment, single cells with no neighbours dominate, making up over 50% of cells present. As the seeding density increases the amount of single cells decreases, with the number of cells in pairs, triples and larger groups increasing (each proportion will always tend to zero as a wider variety of cluster sizes appear). This gives an indication of the seeding densities required to ensure a large proportion of the colonies originate from a single founder cell.

Following the population growth model we can also estimate the time to completely fill a growth area at different seeding densities, Figure 4.13. A more interesting time to consider is the average time that growing colonies first merge due to physical proximity at different seeding densities, Figure 4.14(b). This time, τ, has a linear relationship with $\tau \approx 100 - 0.007 n_0$, allowing us to predict this time for any seeding density. This statistical estimation of τ suggests the latest possible observation time to catch colonies for re-plating before they merge.

Although developed here for a set of particular hESC experiments, the modelling framework can be applied more generally to other cell lines and hiPSCs, and experimental conditions to predict the critical timescales. These results can be used to inform cell seeding density choices to maximise clonal colonies and avoid those arising from more than one founder cell. Future work should explore the role of migrating cells and colonies of clonality and apply the model to hiPSCs.

Chapter 5

Characterising pluripotency

Pluripotency is the defining property of human pluripotent stem cells allowing them to differentiate into any somatic cell in the human body. For the promising clinical applications of hPSCs, improved regulation of pluripotency and differentiation trajectories of their colonies is required. The pluripotency transcription factors which regulate pluripotency are inherently stochastic, inherited asymmetrically upon cell divisions, and more similar in closely related (separated by fewer division events) cells.

In this chapter we use experimental data to quantify the temporal dynamics of the pluripotency transcription factor OCT4 over a cell lifetime. We characterise the internal self-regulation of OCT4, quantify the intra-cellular fluctuations and consider the diffusive nature of its evolution for individual cells and pairs of their descendants. These results are also presented in [187].

5.1 Pluripotency regulation

Human pluripotent stem cells (hPSCs) have the ability to differentiate into all somatic cell types in the human body: the *pluripotency* property. The pluripotency of hPSCs is their defining characteristic, central to their touted applications in drug discovery, regenerative and personalised medicine [12, 13, 16, 237–239]. These promising clinical applications of hPSCs require tight control over colony pluripotency, homogeneity and differentiation trajectories *in-vitro* [51], yet this remains challenging.

The control and optimisation of stem cell pluripotency across colonies is difficult due to the complex inter-regulatory dynamics of pluripotency. At the single-cell level, pluripotency is inherently stochastic. It is suggested that pluripotency is not well defined at the single-cell level but is instead a statistical property of a cell population [60, 143, 144, 240]. Cell pluripotency is also affected by many factors: the local environment [57, 241], interactions with neighbours [59, 61], the cell cycle [27] and the substrate [48]. On the colony scale, complex collective effects of pluripotency emerge. In the presence of restrictive geometries (e.g., the growth area boundaries), differentiated cells form bands of constant width around colony edges as shown in Figure 5.1 [61, 63].

Pluripotency maintenance relies on the inter-regulation of pluripotency transcription factors (PTFs). A network diagram of the regulatory system is shown in Figure 1.11. The key pluripotent gene markers are OCT4, SOX2 and NANOG [242–244]. After several divisions, PTF fluctuations lead to the establishment of sub-populations with varying pluripotency [242]. The differentiation of a stem cell into a specialised cell is the departure

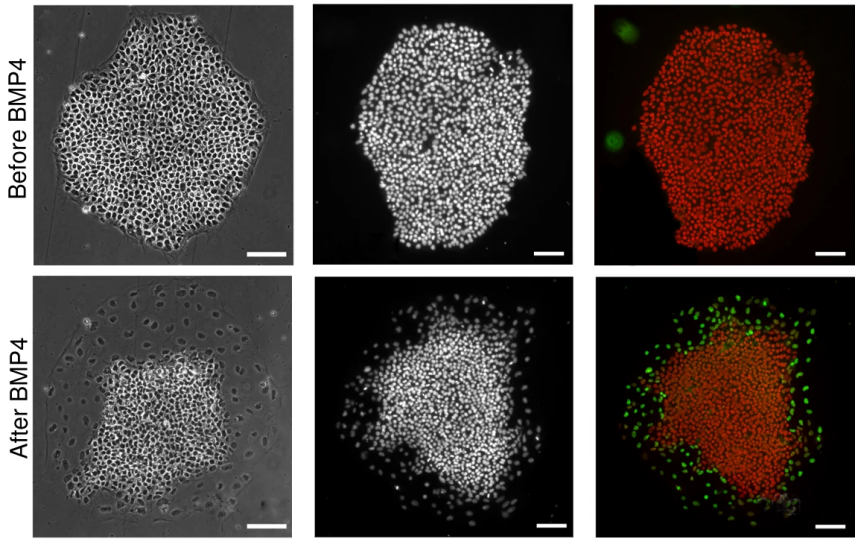

Figure 5.1: Microscopy images of a hPSC colony before and after BMP4 (differentiation inducing agent) showing differentiation around the colony edge. The first column shows phase contrast images, with the middle and final columns showing immunostaining (SOX2 and AP2α) images. In the final panels green highlights the differentiated cells (with low SOX2 and high AP2α) and red the stem cells (with high SOX2 and low AP2α). The scale bar represents 100 μm. Figure adapted from [61].

from the pluripotent state led by PTF destabilisation and their interaction with chemical signalling pathways [42, 43, 242]. This decision of a stem cell to either remain pluripotent or to differentiate is known as its fate decision and is illustrated in Figure 5.2. In many *in-vitro* experiments the differentiation of hPSC populations is biologically induced and facilitated using a differentiation agent, such as BMP4 [245, 246].

It is unknown how much cell fate decisions are led by inherited factors, as opposed to environmental factors and intra-cellular signalling as even clonal (genetically identical) cells under the same conditions make different fate decisions [139]. Colonies exhibit heterogeneous sub-populations of cells with differing levels of PTF expression [42, 242, 247] which suggests a play-off between the potentially disruptive single-cell factors and regulatory community effects [60, 61, 63, 240].

Several experiments have been performed to quantify the behaviour and joint influence of each PTF in the pluripotent cell [40, 248–251]. Their results indicate that PTF expression is highly variable both at the single-cell (time) and colony-level (space) and are subject to intrinsic noise due to interactions at the molecular level and randomness present in the extra-cellular environment [144, 252]. Thus, heterogeneity and stochasticity are inherent properties of pluripotent stem cell populations [144, 240, 253] that hinder their clonal expansion in culture [254–257].

A narrow range of PTF abundance is necessary for maintained pluripotency [258, 259]

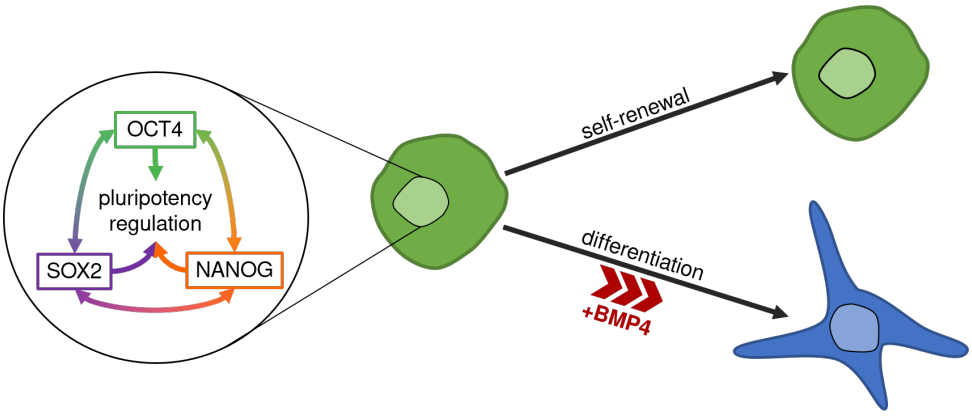

Figure 5.2: Stem cells can either sustain pluripotency and self-renew through mitosis, or differentiate into specialised cells. Pluripotency is regulated by a complex network of transcription factors, including OCT4, NANOG and SOX2. Differentiation *in-vitro* can be induced by the addition of a differentiation agent, such as BMP4.

and even weak fluctuations can bias cell fate decisions in the G1 phase of the cell cycle [260]. Furthermore, the PTFs are inherited asymmetrically as a cell divides, biasing the fate of the daughter cells and contributing to colony heterogeneity [41, 261, 262] with the decision to differentiate largely determined before any differentiation stimulus is introduced [41]. A quantification of the stochastic, temporal dynamics of PTFs is necessary to examine the knock-on effects to cell fate.

Recent mathematical models of these processes focus on describing pluripotency and cell fate decisions to guide the optimisation and control of pluripotency *in-vitro* (discussed in Chapter 1 [85, 92]) and are informed by recent studies of fluctuations of PTFs throughout colonies [41, 240, 260] and the spatial patterning of differentiation [61, 63]. Many models use coupled stochastic differential equations to describe PTF fluctuations [140, 147, 150] while others use a gene network analysis framework [94, 156] or take a mechanistic approach [157].

Although the dynamics of OCT4 are complex, affected by many genetic factors and closely regulated by the other PTFs [43, 242, 263], here we aim to isolate autonomous properties of OCT4 to facilitate the development of descriptive mathematical models. This quantification of OCT4 will provide a basis for identifying systematic similarities and differences between PTFs in future experiments, and highlights some key indicators of cell fate. As our quantitative understanding of PTF regulation increases, more complex regulatory properties can be considered to build fundamental models.

In the next sections we use the experimental data from [41] of time-lapse fluorescent measurements of the OCT4 levels in cells in a growing hESC colony to quantify the dynamics of intra-cellular OCT4.

5.2 Characterising temporal OCT4

In this section we use the experimental data from [41] of time-lapse OCT4 expression in cells in a growing hESC colony to quantify the dynamics of intra-cellular OCT4. We describe quantitatively the fluctuations in OCT4 in relation to cell fate and the addition of the differentiation agent BMP4. A deeper insight into the spatial distribution of OCT4 in the colony and cell fate transitions, based on the same experiment is given in [187]. This quantitative analysis, along with [41] and [187], provides the basis for the developments in mathematical and statistical models of pluripotency presented in Chapter 6.

Experiment 3 This experiment was carried out by Purvis Lab (University of North Carolina, School of Medicine) and is published in [41].

The OCT4 levels (mean OCT4-mCherry fluorescence intensity) in a human embryonic stem cell colony were determined and cells were live-imaged for approximately 70 hours. At 40 hours the differentiation agent BMP4 was added to the culture. The cell IDs, ancestries and positions were extracted along with their OCT4 immuno-fluorescence intensity values (reported in arbitrary fluorescence units, a.f.u.). Each cell was classified according to its final fate status as either pluripotent, differentiated or unknown using expression levels of CDX2. Full experimental details are given in Appendix A and [41]. The experimental data set is available in [41].

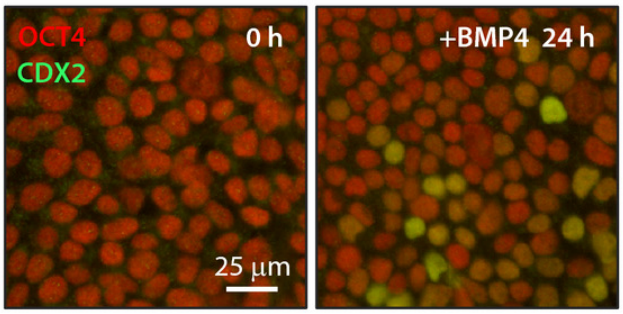

Figure 5.3: Microscopy images of the hESC colony from Experiment 3. Red staining shows the OCT4 levels and green the CDX2 levels (which indicates the presence of differentiated cells). Figure adapted from [41].

5.2.1 Colony growth summary

The colony begins from 30 cells and grows over 68 hours (817 time frames) to 381 cells, with 1274 cell cycles elapsing within this time. The differentiation agent BMP4 was added to the cells at 40 hours. The cells are categorised according to their final cell fate status as pluripotent, differentiated, or unknown (could not confidently be assigned as pluripotent or differentiated). The number of cells, N, considered in each cell fate category, pre- and post-BMP4 is shown in Table 5.1.

Snapshots of the colony at times $T = 0\,\text{h}$, $T = 20\,\text{h}$, $T = 40\,\text{h}$ (the time BMP4 is added) and $T = 68\,\text{h}$ (the final recorded time) colour-coded by OCT4 intensity are shown in Figure 5.4. There is clear spatial patterning of cell fates within the colony, with clustering of pluripotent cells in the centre and differentiated cells around the top edge of the colony. Spatial analysis shows that this patterning begins emerging at around $T = 20$ hours (20 hours before BMP4 addition) for differentiated cells, and at around $T = 50$ hours (10 hours post BMP4 addition) for pluripotent cells [187]. Although here we focus on quantifying the temporal regulation in OCT4, we must keep in mind that there is a spatial correlation between the cell fates.

N	Pre-BMP4	Post-BMP4	All times
Pluripotent	96	422	518
Differentiated	22	111	133
Unknown	112	511	623
All fates	230	1044	1274

Table 5.1: The number of cells, N, in each of the cell fate and pre- and post-BMP4 categories. A post-BMP4 is any cell present at 40 hours or later. There are 1274 cells in total.

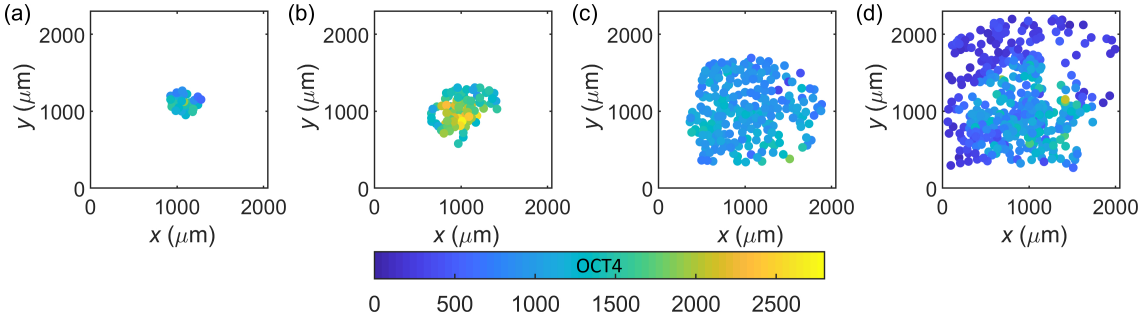

Figure 5.4: Snapshots of the colony at (a) $T = 0\,\text{h}$, (b) $T = 20\,\text{h}$, (c) $T = 40\,\text{h}$ (at the addition of BMP4) and (d) $T = 68\,\text{h}$ (final time). The cells are coloured according to their OCT4 intensity levels. Note that the circles are not indicative of cell or nucleus size.

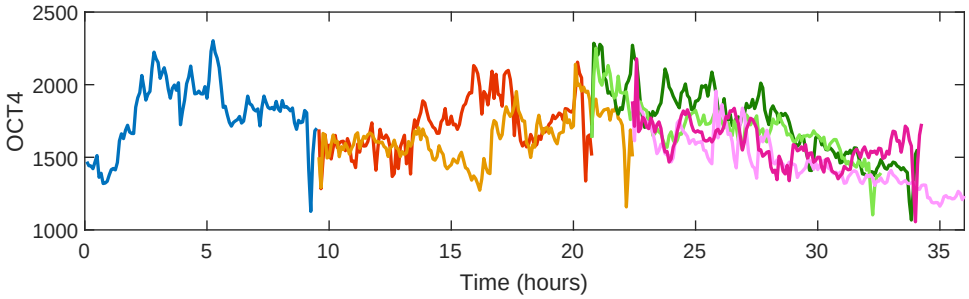

Figure 5.5: Example OCT4 time series from a cell at the start of the experiment (cell ID 25, blue) and its descendent cells.

For every cell in the colony the corresponding time series is available for the abundance of OCT4 within the cell during its lifetime: OCT4(t_1), OCT4(t_2), ..., OCT4(t_n), where $t_1 = 0$ and t_n are the start and end of the cell cycle for the cell, respectively. The values of t_n range from 15 minutes to 30 hours across the population. We will use the notation t_i to describe time-steps in terms of the cell cycle and T_i for experimental time (between 0 and 68 hours). The OCT4 time series for a cell at the beginning of the experiment and its descendants are shown in Figure 5.5.

The number of cells in the colony as a function of time, $N(T)$, for all cells and each cell fate is shown in Figure 5.6. The whole colony follows exponential growth, $N(T) = N_0 e^{\lambda T}$, with the growth rate $\lambda = 0.042 \pm 0.0003\,\mathrm{h}^{-1}$ and the initial condition $N_0 = 33$ cells. Note that when estimated from a least-squares fitting the parameter N_0 differs slightly from the actual initial number of cells (30). For differentiated cells, the growth rate is significantly higher in the first 25 hours, so we consider the fit only after this time. The fittings, along with those for the colony split by cell fate are shown on a linear-linear scale in Figure 5.6(a) and a log-linear scale in Figure 5.6(b) with $N_0 = 14, 6$ and 16 cells, and $\lambda = 0.040 \pm 0.00016$, 0.033 ± 0.0003 and $0.043 \pm 0.0007\,\mathrm{h}^{-1}$ for pluripotent, differentiated and unknown cells respectively. There is a visible plateau, deviating from exponential growth just before 60 hours, originating from the unknown cell population. Both pluripotent and differentiated cell populations continue to follow exponential growth for the duration of the experiment. The growth rate for the colony as a whole corresponds to a doubling time of 16 ± 0.01 hours, as noted in [41] and consistent with other reports [33, 82]. The corresponding doubling times for the different cell fates are 17 ± 0.004, 21 ± 0.008 and 16 ± 0.01 hours for pluripotent, differentiated and unknown cells respectively. As expected, the pluripotent cells proliferate significantly faster than the differentiated cells.

In the next sections we consider the correlation and similarities in OCT4 expression between pairs of descendant ('sister') cells, and proceed to characterise how the drift in OCT4 similarity between these cells emerges over cell lifetimes. We analyse the inherent fluctuations in OCT4 expression and quantify its self-regulatory properties using the Hurst

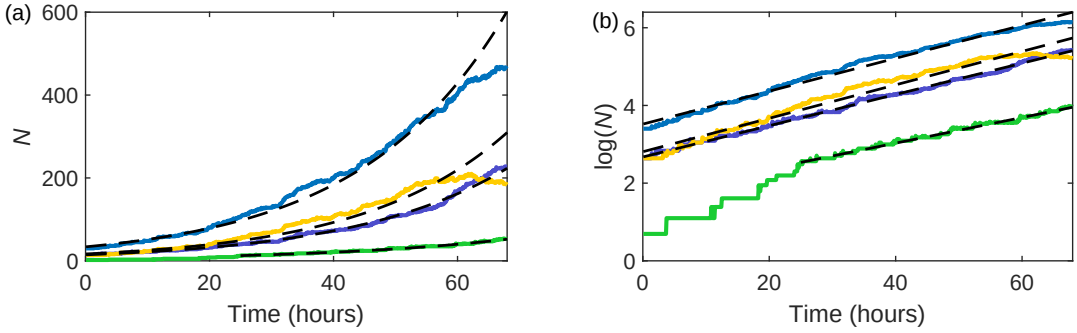

Figure 5.6: (a) The number of all (blue), pluripotent (purple), differentiated (green) and unknown cells (yellow) in the colony over time, $N(T)$, with exponential growth fits $N(T) = N_0 e^{\lambda T}$ (black dashed). The initial estimated conditions N_0 are 33, 14, 5 and 16 cells with the growth rates (doubling times) $\lambda = 0.042 \pm 0.0003$ (16 ± 0.01), 0.040 ± 0.00016 (17 ± 0.004), 0.033 ± 0.0003 (21 ± 0.008) and $0.043 \pm 0.0007\,\mathrm{h}^{-1}$ ($16 \pm 0.01\,\mathrm{h}$) for all, pluripotent, differentiated and unknown cells, respectively. The plateau of unknown cells just before 60 hours causes a deviation from the exponential fit. This is also visible in $N(T)$ for all cells. (b) The corresponding data on a log-linear scale used to obtain the exponential fits.

exponent, autocorrelation and diffusion analysis. Where appropriate, we consider the pluripotent and differentiated cells, pre- and post-BMP4 groups of cells separately to identify any diagnostic factors.

5.2.2 OCT4 in sister cells

The analysis in [41] shows that upon cell division the ratio between the OCT4 values of sister cells is centred around a 1:1 distribution, meaning that although asymmetric pluripotency inheritance does occur (for example, 38% of divisions occur in the ratio 5:6 or more extreme), on average sister cells start with similar levels of OCT4. The OCT4 levels are also more similar in closely related cells, i.e., sister cells and cousin cells show significant similarity when compared with random pairs of cells [41].

We can consider the strength of the correlation in the time series for the OCT4 abundance in sister cells over their whole lifetimes by calculating the Pearson correlation coefficient, ρ. Before calculating the correlation, each OCT4 time series was de-trended (using MATLAB's inbuilt function *detrend* which subtracts the best-fit line from the data) to account for any confounding similarities in sister cells that may be present due to their shared environment. Note that for sister pairs with unequal cell cycle times, the correlation is calculated for the time series of the length up to the minimum cell cycle time. The distribution of ρ for pluripotent and differentiated, and pre- and post-BMP4 sister cells is shown in Figure 5.7(a) and (b). Here it is necessary to pool pre- and post-BMP4 cells together for a cell fate comparison, and vice versa, to keep good

statistics. The mean correlations, $\overline{\rho}$, are given in Table 5.2 and show moderate positive correlations across all categories.

There is no difference in this respect between cells of different fates, both with $\overline{\rho} = 0.5$ (± 0.2 and ± 0.3 for pluripotent and differentiated cells, respectively). The Kolmogorov-Smirnov test provides no evidence at the 95% confidence level to reject the null hypothesis that the distributions of ρ for pluripotent and differentiated cells, shown in Figure 5.7(a), are the same. Sister cells pre-BMP4 show a weaker correlation than those post-BMP4, with $\overline{\rho} = 0.3 \pm 0.2$ and $\overline{\rho} = 0.5 \pm 0.3$, respectively. The Kolmogorov-Smirnov test provides evidence at the 95% confidence level to reject the null hypothesis that the two distributions of Figure 5.7(b) are the same. This suggests that BMP4 treatment exacerbates the similarities in sister cell OCT4 expression. These results quantify the regulation between closely related cells and further illustrate that this regulation is systematic and importantly, still present when confounding external trends are removed.

We can also quantify how this correlation between sister cells varies throughout their lifetimes. The initial and final OCT4 values for all sister cells are shown in Figure 5.7(c) and (d). The initial values follow a very close relationship (as also shown by the OCT4 ratio splitting distribution [41]), with a correlation of $\rho = 0.99$ and the trend line $\mathrm{OCT}_1(t_0) = (1 \pm 0.003)\mathrm{OCT}_2(t_0)$. Note that the labelling of cell 1 and cell 2 is entirely arbitrary, with one sister cell labelled cell 1 and the other cell 2. By the end of their respective lifetimes, the distribution spreads, with a correlation of $\rho = 0.78$ and a line of best fit $\mathrm{OCT}_1(t_n) = (0.97 \pm 0.2)\mathrm{OCT}_2(t_n)$.

In the next section we will consider the behaviour of OCT4 from the initial point of possible asymmetric inheritance to the final time before mitosis at the end of the cell lifetime and characterise how this drift of similarity in sister cells occurs.

$\overline{\rho}$	Pre-BMP4	Post-BMP4	All times
Pluripotent	-	-	$0.5 \pm 0.3 (0.02)$
Differentiated	-	-	$0.5 \pm 0.2\,(0.04)$
All fates	$0.3 \pm 0.2\,(0.03)$	$0.5 \pm 0.3\,(0.01)$	$0.5 \pm 0.3\,(0.01)$

Table 5.2: The mean correlation, $\overline{\rho} \pm$ standard deviation (standard error) between sister cells.

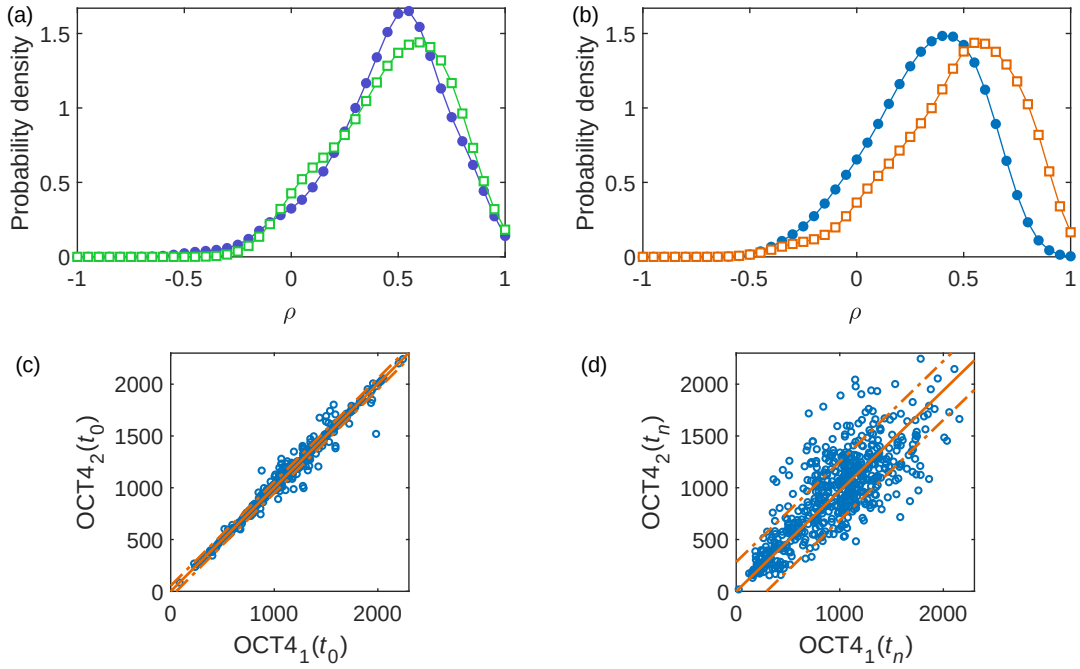

Figure 5.7: The correlation, ρ, between temporal OCT4 in sister cells where both sisters cells were (a) pluripotent (purple filled circles) and differentiated (green unfilled squares) and (b) cells pre-BMP4 (blue filled circles) and post-BMP4 (orange unfilled squares). OCT4 values for all sister pairs (c) at the start and (d) end of their cell cycles. The lines of best fit (orange solid lines) with standard errors in predicting a future observation (dashed lines) are (c) $\mathrm{OCT}_1(t_0) = (1 \pm 0.003)\mathrm{OCT}_2(t_0)$ with $R^2 = 0.98$ and (d) $\mathrm{OCT}_1(t_n) = (0.97 \pm 0.02)\mathrm{OCT}_2(t_n)$ with $R^2 = 0.78$.

5.2.3 Temporal dynamics

In this section we quantify the temporal behaviour of the OCT4 dynamics on the cellular level over the course of a cell lifetime. We consider the variability between discrete time-steps and quantify the self-regulatory behaviour of OCT4 using several methods.

OCT4 distribution

To get an overall view of the OCT4 expression levels, the distributions of all measured OCT4 values for pluripotent and differentiated cells, pre- and post-BMP4 are considered, shown in Figure 5.8. Pre-BMP4, the differentiated cells show a skewed distribution, with a higher preference of lower OCT4 expressions than the pluripotent cells. This is fitting with the fact that the analysis in [41] suggests the tendency to differentiate is largely pre-determined before the addition of the differentiation stimulus. Post-BMP4, both pluripotent and differentiated cells also show a reduction in their OCT4 expression, with the effect seen more strongly in the differentiated cells. It is expected that the BMP4 causes a reduction in OCT4 in the differentiated cells, but it is interesting that the same effect (to a lesser extent) is also present in the cells which remain pluripotent. The average (median) OCT4 expression levels, $\overline{\text{OCT4}}$ are shown in Table 5.3.

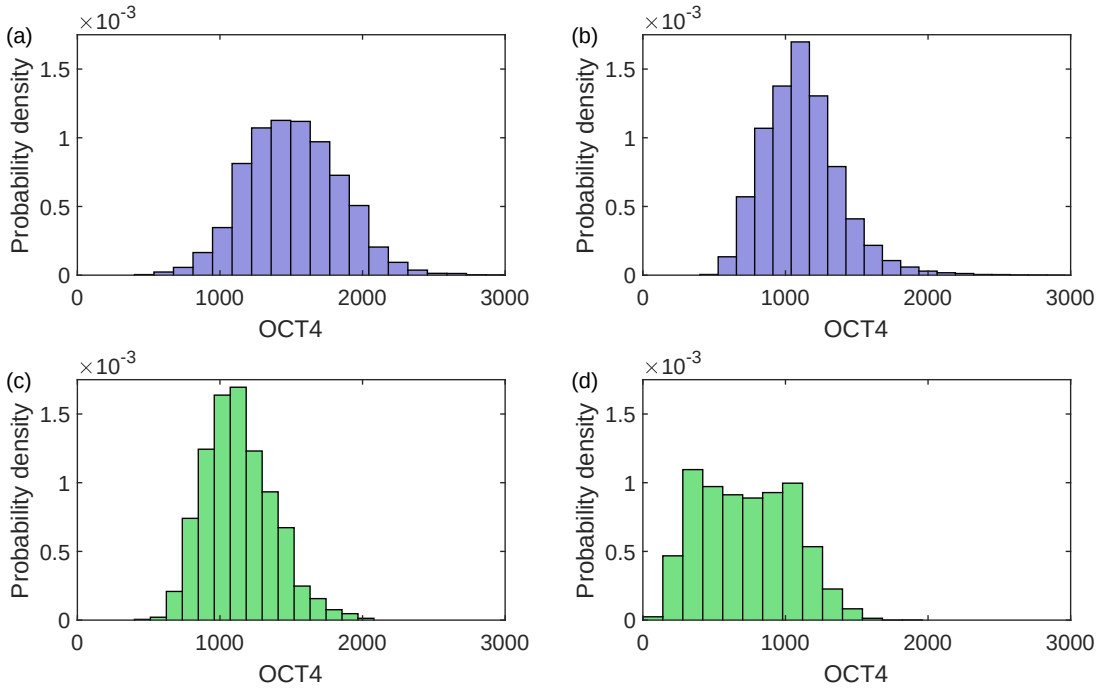

Figure 5.8: The distributions of OCT4 for pluripotent cells (purple) (a) pre-BMP4 and (b) post-BMP4, and differentiated cells (green) (c) pre-BMP4 and (d) post-BMP4.

$\overline{\text{OCT4}}$	Pre-BMP4	Post-BMP4	All times
Pluripotent	1500 [1280 1730]	1090 [930 1260]	1160 [980 1380]
Differentiated	1100 [960 1290]	720 [450 990]	840 [510 1070]
All fates	1420 [1190 1670]	1050 [850 1230]	1110 [900 1320]

Table 5.3: The median OCT4 values, with lower and upper quartiles given as $\overline{\text{OCT4}}$ [lower quartile upper quartile].

Variability at short time scales

Even small fluctuations in PTF abundance impact cell fate [260] with both high and low PTF values resulting in differentiation [258, 259]. The mathematical quantification of PTF fluctuation will facilitate the description of pluripotency over discrete time-steps, fitting for time-lapse experiments such as the one considered here [41]. First, we consider the change in the intra-cellular OCT4 abundance between the five minute intervals, $t_1, t_2, ...t_n$, as $\Delta\text{OCT4} = \text{OCT4}(t_i) - \text{OCT4}(t_{i-1})$. Our choice of the five minute intervals allows for good statistics with 90% of cells having greater than 50 data points. It is likely that a large proportion of these individual fluctuations will be due to experimental noise, but considering all of these values together reveals the average behaviour.

The distributions of ΔOCT4 for pluripotent and differentiated cells, pre- and post-BMP4 are shown in Figure 5.9. For both fates across all times, the change in OCT4 is centred around zero (although the individual values range from -1300 to 1200). This means that, on average, the change in OCT4 is bi-directional for both pluripotent and differentiated cells. There is no preference for the OCT4 abundance to increase or decrease in a time-step, the fluctuations are symmetric overall. Interestingly, although symmetric, the distributions are not Normal (confirmed by the Kolmogorov-Smirnov test at the 95% confidence level) due to a narrower and steeper peak, visible in Figure 5.9. A Laplace distribution, Laplace$(\mu^\dagger, b)$, with probability density function $\mathrm{e}^{-|x-\mu^\dagger|/b}/2b$, better fits the experimental data in all cases. The parameters can be estimated using the maximum likelihood estimators $\hat{\mu^\dagger}$ and $\hat{b}$, where $\hat{\mu^\dagger}$ is the sample median and $\hat{b}$ is the mean absolute deviation from the median

$$\hat{b} = \frac{1}{N}\sum_{i-1}^{N} |x_i - \hat{\mu^\dagger}|. \tag{5.1}$$

These estimated parameters for Laplace fittings to ΔOCT4 are given in Table 5.4.

Post-BMP4 addition, the distributions for both cell fates become significantly narrower, with the parameter b showing a reduction of 49% for differentiated cells, and 33% for the pluripotent cells. There is also a subtle skew in the differentiated cells towards negative values of ΔOCT4 which is consistent with the fact that the OCT4 levels overall decrease

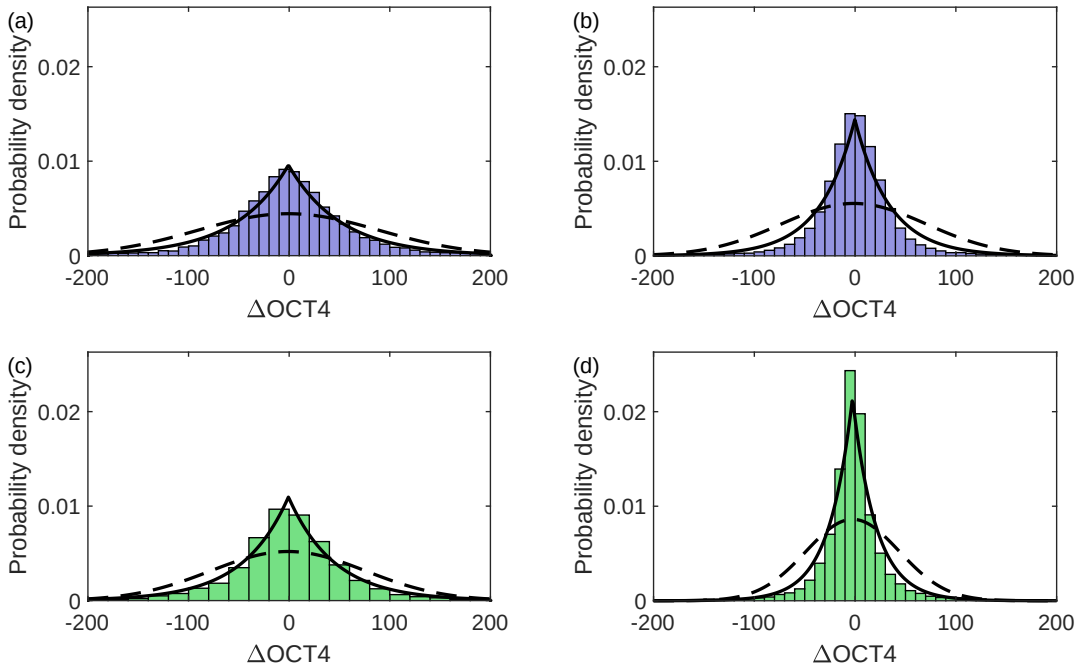

Figure 5.9: Distributions of the change in OCT4 between the five minute time frames (ΔOCT4) for pluripotent cells (a) pre-BMP4 and (b) post-BMP4, and differentiated cells (c) pre-BMP4 and (d) post-BMP4. Solid lines show the Laplace distribution fittings, Laplace($\mu^\dagger, b$), with the parameters (a) $\mu^\dagger = -0.7$ and $b = 52.1$, (b) $\mu^\dagger = -0.3$ and $b = 34.7$, (c) $\mu^\dagger = -0.7$ and $b = 45.6$ and (d) $\mu^\dagger = -2.6$ and $b = 23.3$. Dashed lines show the Normal distribution fittings.

after the BMP4 addition. The narrowing of the distributions is due to a preference of smaller changes in OCT4 in all cell fates provoked by the differentiation agent. This could be driven by induced selectivity caused by the BMP4 addition (i.e., the BMP4 causes a systematic change, producing a preference for smaller ΔOCT4 values), or it could suggest some collective self-regulation [60]. Further experiments (varying colony size) are needed to investigate if this is a collective behaviour effect. It is expected, since the differentiated cells are most affected by the BMP4, that this group would show the biggest reduction in variation and therefore the strongest regulation in their OCT4 values.

We can also consider the self-similarity of the OCT4 series using the Poincaré map [264, 265]. Poincaré maps offer an efficient diagnostic for analysing the self-similarity of a series. A timed signal is plotted against itself after a time delay (here the time delay is 5 min) and its scatter pattern reflects the randomness and variability of the dynamics, thus giving a representation of the correlation between consecutive values of the time series. For each cell, its OCT4 time series can be plotted against itself with one time-step delay, i.e., OCT4(t_i) against OCT4(t_{i+1}), as shown in Figure 5.10. In the Poincaré plots we

Laplace($\mu^{\dagger}$,b)	Pre-BMP4	Post-BMP4	All times
Pluripotent	(−0.7, 52.1)	(−0.3, 34.7)	(−0.3, 38.5)
Differentiated	(−0.7, 45.6)	(−2.6, 23.3)	(−2.4, 28.1)
All fates	(−0.8, 50.8)	(−0.9, 32.4)	(−2.4, 28.1)

Table 5.4: The parameters from the Laplace($\mu^{\dagger}$,b) fittings to the ΔOCT4 distributions shown in Figure 5.9.

SD1, SD2	Pre-BMP4		Post-BMP4		All times	
	SD1	SD2	SD1	SD2	SD1	SD2
Pluripotent	64	1430	51	1150	54	1320
Differentiated	54	1070	33	990	38	1090
All fates	62	1440	48	1170	51	1330

Table 5.5: Quantitative results for the Poincaré map ellipse fittings. The major axis (SD1) and minor axis (SD2) from fitting ellipses to the plots shown in Figure 5.10.

colour the values according to their normalised frequency. The data (changes in OCT4) is binned into 100 groups before the relative frequency of that group (a value between 0 and 1) is calculated.

By assessing qualitatively the shape formed by the return maps shown in Figure 5.10, we observe changes in the distribution of points between pluripotent and differentiated cells, pre- and post-BMP4. Even pre-BMP4 addition, the differentiated cells show less variation compared to the pluripotent cells, with the addition of BMP4 exacerbating this effect. Quantitatively these results can be described by fitting an ellipse to the shape formed by the data plots and measuring the dispersion along the major SD1 and minor SD2 axes, given in Table 5.5.

This information quantifies step changes in OCT4, suggesting the use of the Laplace distribution to simulate variation and shows that the addition of BMP4 provokes tighter self-regulation across both cell fates. It also highlights that even between small time increments such as these, the fluctuations post BMP4 should be considered separately for cells of different fates, not only in terms of their average, as expected, but also their variability. Note that this allows us to capture the nature of the variation in OCT4 only and further aspects of the behaviour need to be considered to fully describe the OCT4 regulation over time.

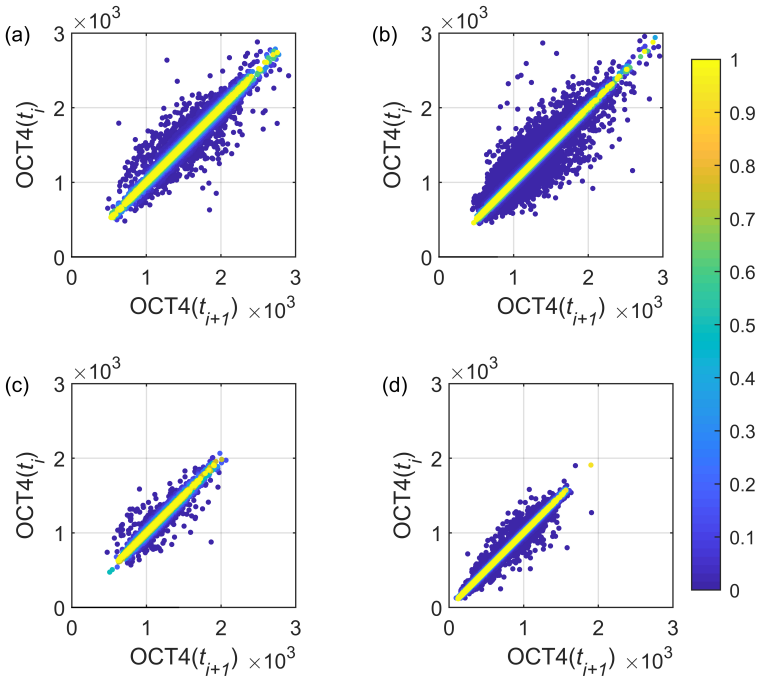

Figure 5.10: Poincaré maps for the OCT4 signal for pluripotent cells (a) pre-BMP4 and (b) post-BMP4, and differentiated cells (c) pre-BMP4 and (d) post-BMP4. The colour bar shows the normalised relative frequency of the points.

OCT4 self-regulation

To investigate the self-regulation and internal memory of OCT4 during a cell cycle, we consider three related approaches, the Hurst exponent, the autocorrelation function and diffusion analysis.

The Hurst exponent

The Hurst exponent, $0 < H < 1$, is a measure of the memory, or the self-similarity properties of a time series [103, 266]. It is defined as

$$E\left[\frac{R(n)}{S(n)}\right] = Cn^H \text{ as } n \to \infty, \tag{5.2}$$

with $R(n)$ the range of the first n cumulative deviations from the mean and $S(n)$ their standard deviation. $E[\cdot]$ denotes the statistical expectation, n is the number of data points in the time series and C is a constant. The quantity R/S is known as the rescaled range and measures how the apparent variability changes with the length of time considered. For a time series $X_1, X_2, ..., X_n$, with mean $m = \frac{1}{n}\sum_{i=1}^{n} X_i$, R_t can be calculated as

$$R_i = \max(Z_1, Z_2, ..., Z_t) - \min(Z_1, Z_2, ..., Z_t), \tag{5.3}$$

where

$$Z_t = \sum_{i=1}^{t} X_i - m \text{ for } t = 1, 2, ..., n. \tag{5.4}$$

We calculated the Hurst exponent using the function *genhurst* (MATLAB) [267]. Further details on the Hurst exponent, other methods of estimation and its relation to fractional Brownian motion can be found in [103, 266, 268, 269].

The Hurst exponent provides a measure of the long term memory of a time series. If a series is Brownian, $H = 0.5$, then the fluctuations are mutually statistically independent, with the variable just as likely to increase as decrease at each time-step. If the series is persistent, $H > 0.5$, then at each time-step the series is more likely to fluctuate in the same direction as in the previous step, i.e., if in the last time-step there was an increase, it is more likely there will be another increase during the next time-step. For anti-persistence, $H < 0.5$, the series is less likely to fluctuate in the same direction as the previous step.

The Hurst exponent was calculated for all cells which live longer than 50 time frames (4.16 hours). The distributions of all H values for pluripotent and differentiated cells, pre- and post-BMP4 are shown in Figure 5.11. The average Hurst exponents, $\overline{H}$, are given in Table 5.6 for each group. In all cases, the Hurst exponents are significantly less than 0.5, showing moderate anti-persistence. This shows the self-regulation of OCT4 on the intra-cellular scale: if the OCT4 value has just increased, it is more likely to next decrease, and vice versa. This is the case across each cell fate group.

Although the means are within errors of one another, the Kolmogorov-Smirnov test rejects the null hypothesis that the distributions of H for pluripotent and differentiated pre-BMP4 cells are the same at the 95% level. There is no significant difference in H before and after the BMP4 addition for both cell fates (confirmed by the Kolmogorov-Smirnov test at the 95% level) suggesting this aspect of the self-regulatory behaviour is inherent to the cells and unchanged by the differentiation stimulus. This quantification via the Hurst exponent is directly transferable to use in fractional Brownian motion modelling [103, 266, 268, 269].

$\overline{H}$	Pre-BMP4	Post-BMP4	All times
Pluripotent	0.37 (0.08 0.008)	0.37 (0.09 0.004)	0.37 (0.09 0.004)
Differentiated	0.42 (0.08 0.02)	0.39 (0.09 0.009)	0.40 (0.09 0.008)
All fates	0.38 (0.08 0.007)	0.38 (0.09 0.004)	0.38 (0.09 0.004)

Table 5.6: The mean Hurst exponent $\overline{H}$ with (standard deviation, standard error) for all cell categories.

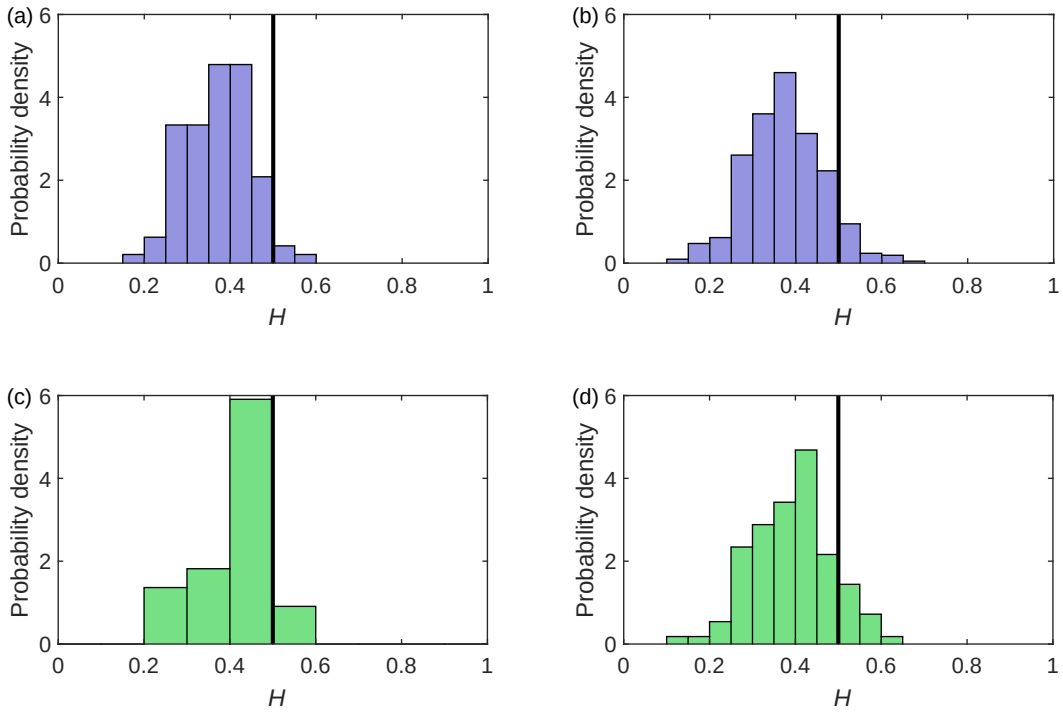

Figure 5.11: The Hurst exponent, H, for pluripotent cells (purple) (a) pre-BMP4 and (b) post-BMP4 and differentiated cells (green) (c) pre-BMP4 and (d) post-BMP4. The black lines show $H = 0.5$, the value for Brownian fluctuations.

Autocorrelation

The anti-persistence can be further explored by considering the autocorrelation of the OCT4 time series. The autocorrelation function, C, is the correlation of a time series with itself at increasing time lags, hence $-1 \leq C \leq 1$ where $C = 0$ signifies no correlation, $C < 0$ a negative correlation (corresponding to anti-persistence) and $C > 0$ a positive correlation (persistence). We calculate the autocorrelations using MATLAB's *autocorr* function (Econometrics Toolbox). The autocorrelation C_i of a time series between x_t and x_{t+i} for time-lag i is given by

$$C_i = \frac{1}{T\sigma} \sum_{t=1}^{T-i} (x_t - \overline{x})(x_{t+k} - \overline{x}), \tag{5.5}$$

where σ is the sample variance of the time series. The decay of the autocorrelation to zero (scaled to cell lifetimes) is presented in the Appendix in [41] and here we extend this to quantify the periods of anti-persistence and consider the periodic nature of the autocorrelation.

Typical autocorrelation functions are shown in Figure 5.12. The majority of the cells follow an autocorrelation similar to the one shown in Figure 5.12(a) (cell ID 46), with initial correlation (inter-dependence between the consecutive measurements) declining to zero, followed by a period of anti-correlation before the autocorrelation settles at zero. There are, however, other behaviours evident. Some cells show several lag intervals of anti-correlation, as in Figure 5.12(c) (cell ID 14), with others showing a positive correlation at a longer time lag before settling at zero, as in Figure 5.12(e) (cell ID 43). The corresponding time series of OCT4 for each example cell are shown in Figure 5.12(b,d) and (f). The intervals of positive correlation are visible as trends in OCT4 (either continued increase or decline), with anti-correlation visible as fluctuations about a horizontal line.

Anti-correlation for a time lag of at least one hour duration is seen in 99% (1255/1274) of cells, and for at least five hours in 86% (1090/1274) of cells. Of the cells with at least one hour anti-correlation visible, 44% show a second period of correlation near the end of their lifetimes (as in Figure 5.12(a) and (c)). Out of these, 65% show one period of anti-correlation (as in Figure 5.12(e)), 31% two periods (as in Figure 5.12(c)), and the remaining 4% three or more. For the 57% of cells with no second period of correlation, 90% of cells show one period, 8% two periods and 2% three or more periods of anti-correlation.

The shortest time lag at which the anti-correlation occurs, t_{AC}, can be calculated for each individual cell. The distribution of t_{AC} for cells with at least one hour anti-correlation is shown Figure 5.13(a) and reveals the critical cell cycle stage in which it first occurs. In all cells with anti-correlation, it has begun by 8 hours into the cell cycle (just over half a cell cycle [41]), suggesting that before they reach the latter halves of their lifetimes the internal self-regulation of OCT4 begins. This could be due to the memory effects or the down-regulation of the PTF which occurs prior to mitosis [270, 271]. The percentage lifetime a cell spends in an anti-correlated state is 40–80% (with a mean of 60%) for both fates, pre- and post-BMP4.

The oscillatory behaviour and decay of the autocorrelation can be captured by the function $C = \cos(2\pi t/a)e^{-t/b}$ [272] (note that this periodicity in the autocorrelation does not necessarily imply periodicity in the time series). These fittings are shown in Figure 5.14 for 25 randomly selected cells in the colony. This quantifies the temporal, periodic decay in the autocorrelation, with the parameter a representing the time-scale of the periodicity, and b the time-scale of the decay (the correlation decay time). Histograms of a and b for all 1274 cells are shown in Figure 5.15. Both distributions are skewed, with medians of 11.7 h and 3.0 h, and 90th percentiles of 30 h and 7 h for a and b respectively. This quantifies the characteristic time-scale of the periodicity and the correlation decay time as less than 30 hours and 7 hours in 90% of cases, respectively.

The correlation time is defined as $\tau = \int_{-\infty}^{\infty} C(t)dt$. We estimate τ as the area under the curve for each autocorrelation function, which produces a mean correlation time across all

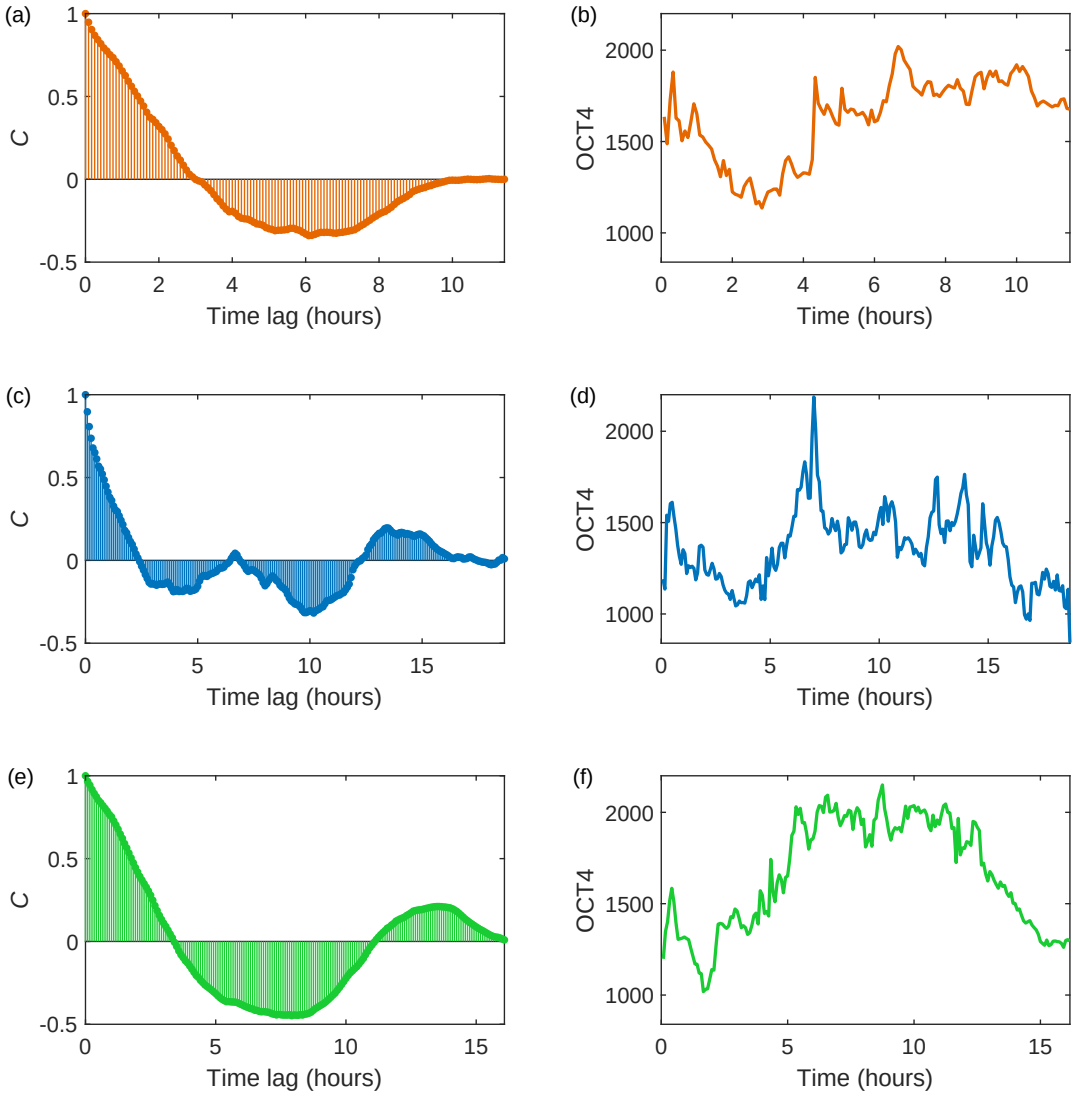

Figure 5.12: Typical autocorrelations showing (a) a period of anti-correlation before settling at zero correlation (seen in 51% of cells), (c) two periods of anti-correlation followed by correlation (28% of cells) and (e) a period of anti-correlation followed by a period of correlation (14% of cells). The panels (b,d) and (f) show the OCT4 variation in time for these cells respectively. The average behaviour is similar to that in (a).

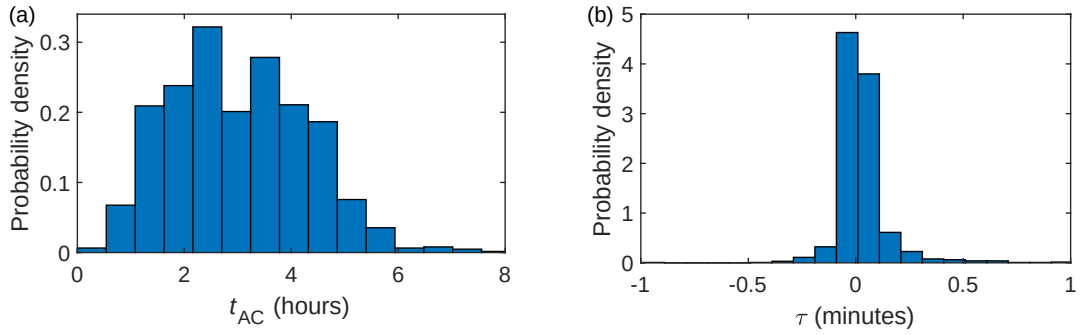

Figure 5.13: (a) The first time at which the autocorrelation becomes negative for all cells, i.e., the time at which the anti-correlation begins within each cell cycle. (b) All correlation times for all cells, $\tau = \int_{-\infty}^{\infty} C(t)dt$.

cells of $\overline{\tau} \approx 0 \pm 0.002\,\mathrm{h}$, indicating that the areas of positive and negative auto-correlations are close to each other. The distribution of all correlation times is shown in Figure 5.13(b).

We can identify the average behaviour by considering all autocorrelations for all cells. The mean (and standard deviation) and median (and interquartile range) autocorrelations $\overline{C}$ for all cells is shown in Figure 5.16(a). Notably the mean and median are comfortably within errors of each other and the autocorrelation is robust to the chosen averaging method. The average autocorrelation decreases to zero at around three hours, followed by a period of negative autocorrelations indicative of anti-persistent behaviour between approximately three and 12 hours. By 13 hours, the average autocorrelation settles at zero, showing no internal memory past this time. These observations are robust to cell fate and the equivalent autocorrelations for pluripotent and differentiated cells are shown in Figure 5.16(b) and (c). This shows that during a cell cycle, there is long-term memory in the OCT4 expression up to around 12 hours, but the nature of the effect differs over this time with initial correlation being replaced by anti-correlation. Notably, the mean autocorrelation is not fully described by $\cos(2\pi t/a)e^{-t/b}$, as the full scale of the anti-correlation is not captured, shown in Figure 5.16(d). These results can be used to inform chemo-dynamical models of the OCT4 production within a cell, with the time scales identified likely to reflect the chemical reaction rates involved.

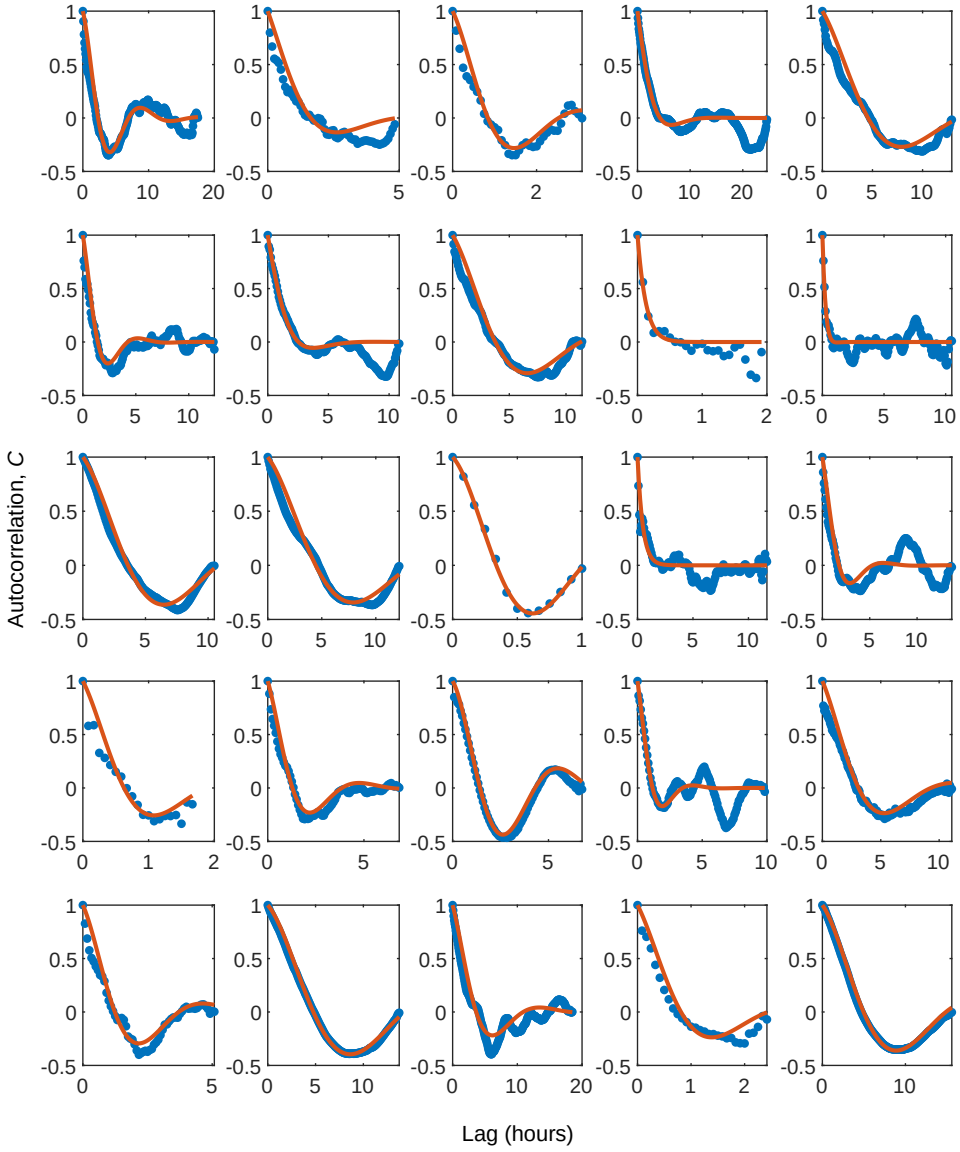

Figure 5.14: Autocorrelation functions for 25 random cells (blue circles) with the fittings $C = \cos(2\pi t/a)e^{-t/b}$ (orange solid lines).

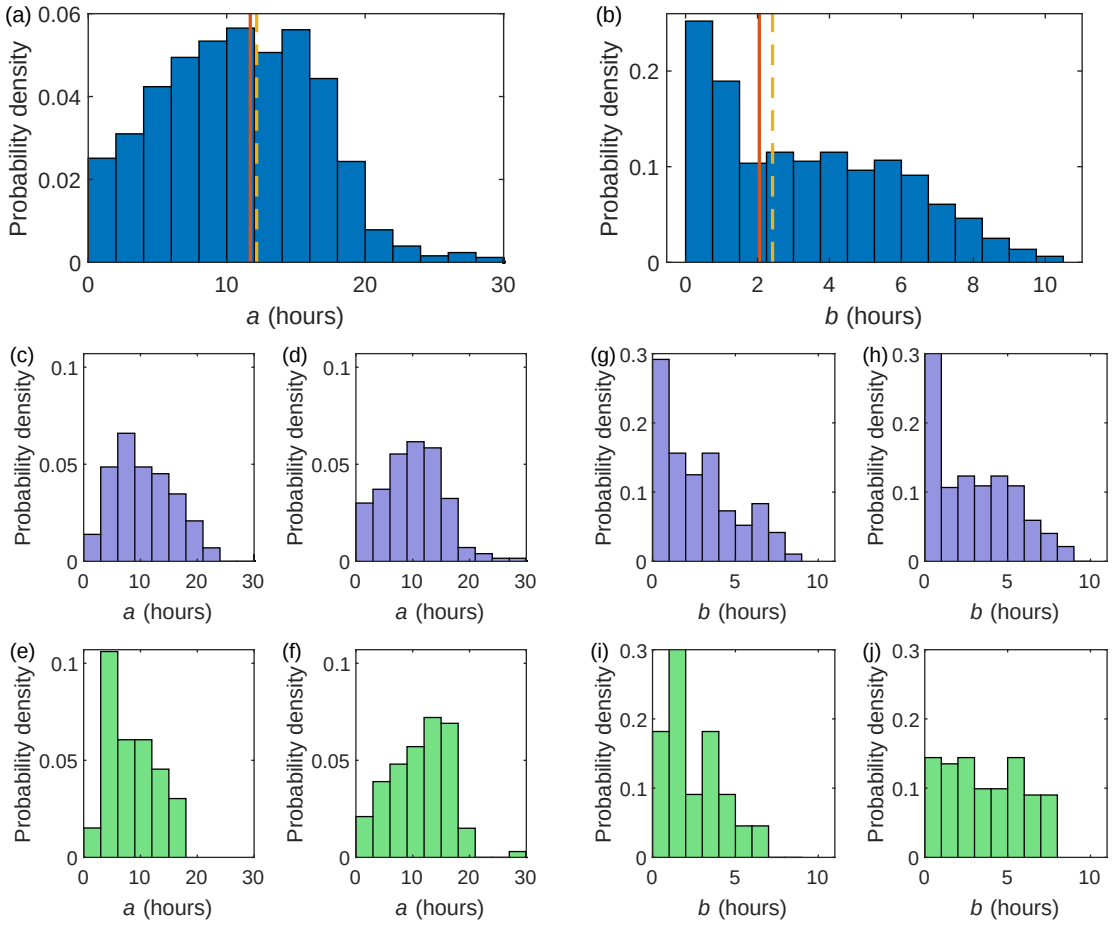

Figure 5.15: The distributions of the parameters (a) a and (b) b from all $C = \cos(2\pi t/a)e^{-t/b}$ autocorrelation fittings. The parameter estimates using the mean and median autocorrelations for all cells are $a = 11.7 \pm 0.92$ and 12.16 ± 0.69 and $b = 2.05 \pm 0.23$ and 2.42 ± 0.21 and are shown as a red solid and yellow dashed line, respectively. The distribution of a and b for pluripotent cells (purple) (c,g) pre-BMP4 and (d,h) post-BMP4 and differentiated cells (green) (e,f) pre-BMP4 and (i,j) post-BMP4.

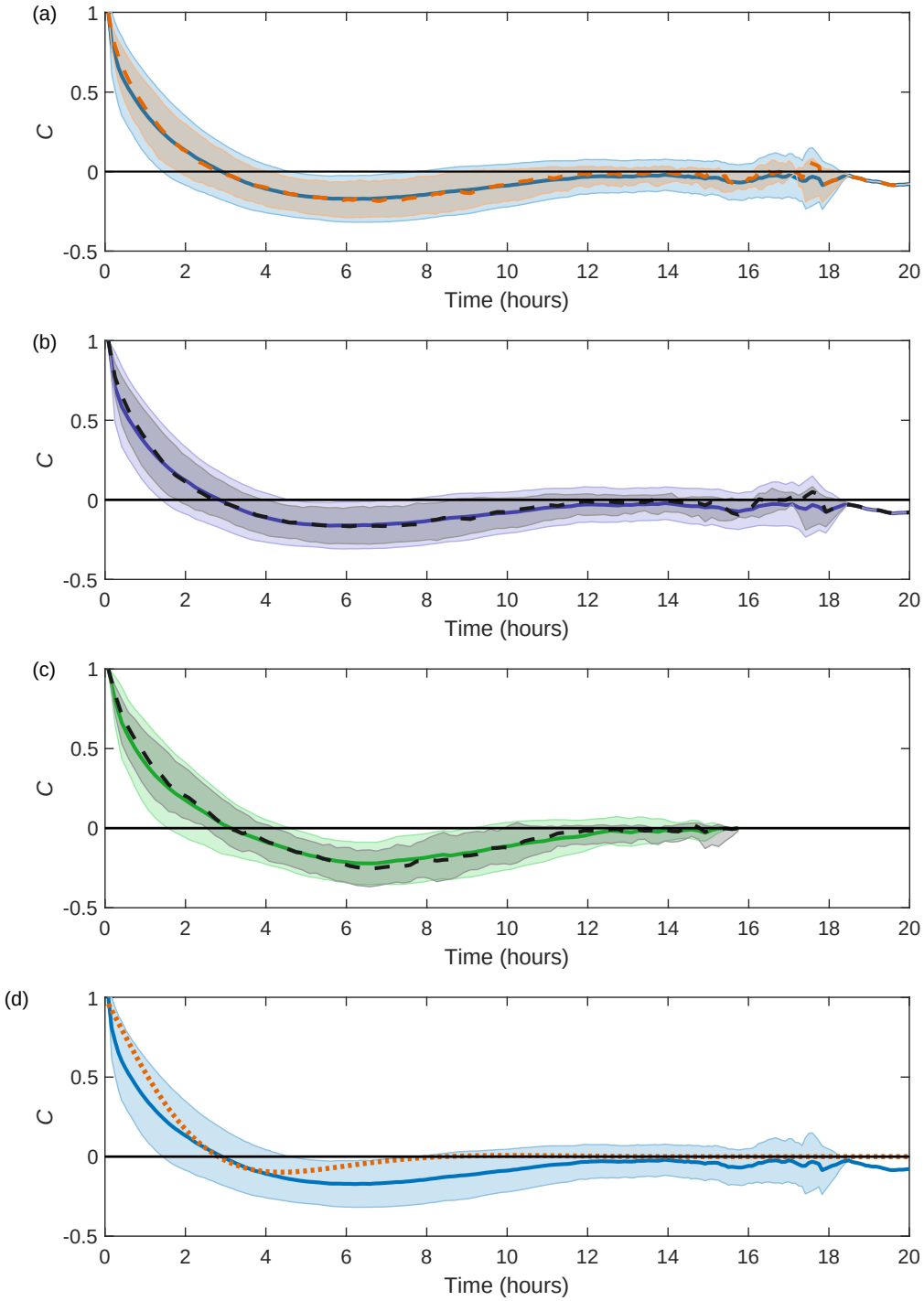

Figure 5.16: (a) The mean (blue solid with standard deviation error bars) and median (orange dashed, with interquartile range error bars) autocorrelation for all cells with increasing time lag. The corresponding autocorrelations for (b) pluripotent (mean purple solid, median black dashed) and (c) differentiated (mean green solid, median black dashed). (d) The mean autocorrelation for all cells with the fitting $C = \cos(2\pi t/a)e^{-t/b}$ (orange dashed), with $a = 10.9 \pm 0.9$ and 11.3 ± 0.8,

Diffusivity

A further method of quantifying the internal regulation of OCT4 is to consider the diffusive behaviour of the time series. The theory of diffusion and random walks, introduced in Chapter 1, is widely used across many biological applications, including stem cells and so it is useful to quantify the OCT4 behaviour within this framework [74, 87–89, 109, 112].

After the cell division with asymmetric inheritance of OCT4 there is a short period of increased fluctuations [41]. Here we consider each OCT4 time series from half an hour after cell division. Each cell has an OCT4 value at the start of its lifetime, denoted OCT_0. The mean square difference of OCT4 over time, $\mathrm{MSD}(t)$, can be calculated as $\langle|\mathrm{OCT4}(t) - \mathrm{OCT4}_0|^2\rangle$, where the angular brackets denote the average across all cells in the group considered. The MSD for all, pluripotent, and differentiated cells, pre- and post-BMP4 between 0 and 15 hours is shown in Figure 5.17. For pluripotent cells, both pre- and post-BMP4, the distinct sub-diffusive behaviour of the MSD is visible, with $\mathrm{MSD} = \beta t^{\alpha}$, $\alpha < 1$. The parameters α and β are shown in Table 5.7. The differentiated cells show diffusive behaviour at early times, with $\alpha \approx 1$, but the limiting of the MSD can still be seen from around 2 hours pre-BMP4 and from 9 hours post-BMP4.

This sub-diffusivity is consistent with the anti-persistence revealed using the Hurst exponent and autocorrelation. In agreement with those results, on average, the intracellular OCT4 abundance behaves in a sub-diffusive manner throughout a cell lifetime. This has a knock-on effect for the relationship between sister cell OCT4 which is considered below.

This further quantifies the self-regulatory behaviour of OCT4 within the diffusion framework, a fundamental starting point for many mathematical models. The anti-persistence of OCT4 suggests possibilities for mathematical modelling methods to capture the internal regulation of pluripotency, including fractional Brownian motion and correlated random walk theory.

	Pre-BMP4		Post-BMP4	
	α	β	α	β
Pluripotent	0.59 ± 0.03	42000 ± 2700	0.54 ± 0.03	35000 ± 2300
Differentiated	0.88 ± 0.13	54000 ± 3300	1.04 ± 0.04	13000 ± 900
All fates	0.52 ± 0.03	44000 ± 2500	0.70 ± 0.02	26000 ± 2700

Table 5.7: Parameters for $\mathrm{MSD} = \beta t^{\alpha}$ fittings.

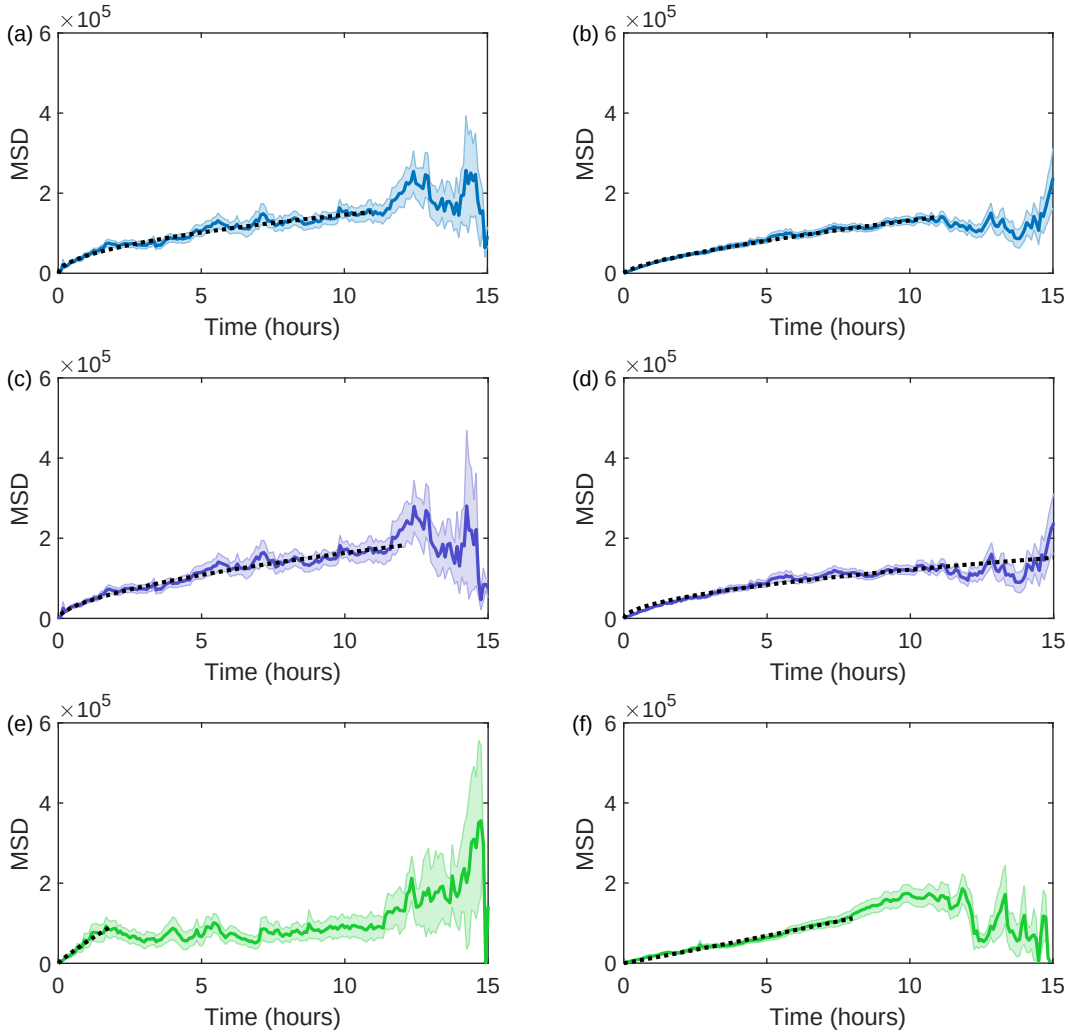

Figure 5.17: The MSD for all cells (a) pre and (b) post-BMP4, pluripotent cells (c) pre- and (d) post-BMP4 and differentiated cells (e) pre- and (f) post-BMP4. The black dashed lines show the fits MSD $= \beta t^{\alpha}$ with the parameters (a) $\alpha = 0.52 \pm 0.03$, $\beta = 44000 \pm 2500$, (b) $\alpha = 0.70 \pm 0.02$, $\beta = 26000 \pm 1000$, (c) $\alpha = 0.59 \pm 0.03$, $\beta = 42000 \pm 2700$, (d) $\alpha = 0.54 \pm 0.03$, $\beta = 35000 \pm 2300$, (e) $\alpha = 0.88 \pm 0.13$, $\beta = 54000 \pm 3300$ and (f) $\alpha = 1.0 \pm 0.04$, $\beta = 13000 \pm 900$.

Diffusivity between pairs

The drift in OCT4 values between sister cells can conveniently be considered within the diffusion framework. In the previous section we showed that intra-cellular OCT4 abundance behaves in a sub-diffusive manner throughout a cell lifetime. We can examine the effect this has on pairs of sister cells and the implications for modelling.

Consider the difference between the OCT4 levels in sister cells through time, $\mathrm{OCT}_1(t)$ and $\mathrm{OCT}_2(t)$, denoted as $\delta(t) = |\mathrm{OCT}_1(t) - \mathrm{OCT}_2(t)|$. Note that the labelling of cell 1 and cell 2 is entirely arbitrary. We can consider the mean square difference for δ, $\mathrm{MSD} = \langle|\delta - \delta(0)|^2\rangle$, where angular brackets denote the average across all sister pairs. The MSD is shown in Figure 5.18 and reveals the sub-diffusive behaviour; a result of the sub-diffusive nature of individual cells.

We can compare the sister cell pairs to a Brownian diffusive simulation, shown in Figure 5.19, illustrating realisations for three pairs of cells and their corresponding δ, along with δ for all 564 experimental pairs, and 500 simulated pairs. This leads to the MSD behaviour shown in Figure 5.19(g) and (h), with the experimental pairs showing a suppressed MSD behaviour in comparison to a Brownian diffusive simulation.

This highlights that although the difference between OCT4 levels in sister cells does increase stochastically over the cell lifetimes, it does so in a suppressed, sub-diffusive manner. These results further quantifies similarities between sister cells and their diffusive behaviour, and suggests extensions to Brownian motion for modelling purposes, such as fractional or geometric Brownian motion.

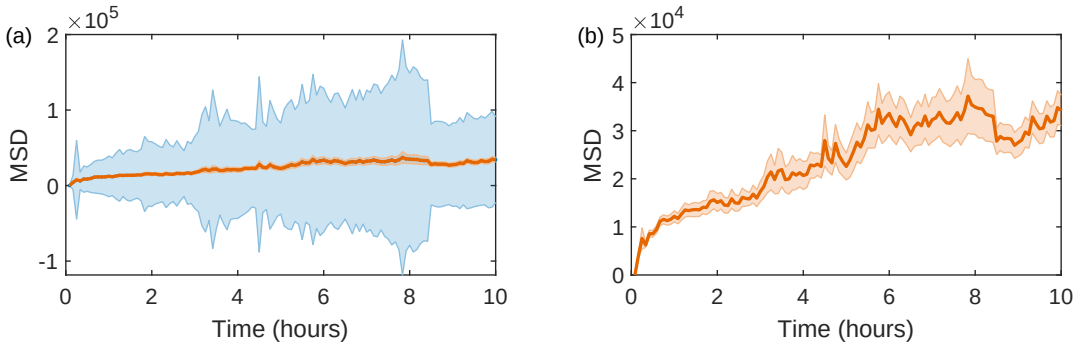

Figure 5.18: (a) The MSD, $\langle|\delta-\delta(0)|^2\rangle$ in OCT4 values between sister cells with standard error (orange) and standard deviation (blue) errors. (b) The MSD with standard error bars only.

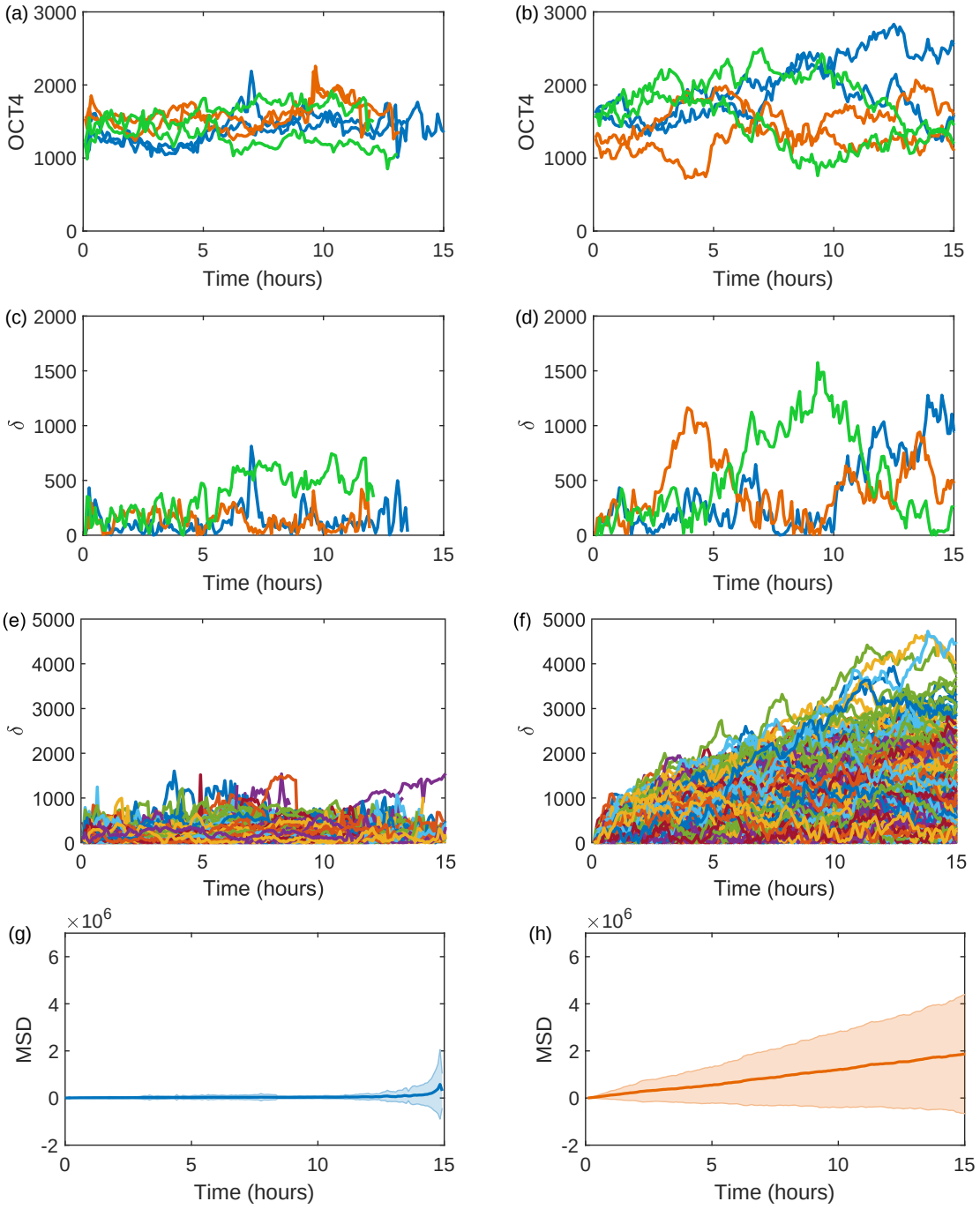

Figure 5.19: A comparison of the behaviour of OCT4 within experimental sister pairs (left) and a simulated pairs (right) which behave diffusively with $\sigma = 80$ to correspond to the experimental data. (a,b) A realisation of three pairs with (c,d) the difference in OCT4 between these pairs over time, δ, and (e,f) δ for all 564 experimental pairs, and 500 simulated pairs. The MSD for (g) experimental and (h) simulated sister pairs of cells with standard deviation error bars.

OCT4 decline prior to cell division

A sharp decline in OCT4 levels occurring before cell division is noted in [41], in keeping with the transcription factor down regulation known to occur before mitosis [270, 271]. This phenomena can be quantified, with the decrease in OCT4 beginning, on average, 35 minutes (0.58 hours) before cell division, lasting for 15 minutes (0.25 hours), and showing a reduction of 22% from the interphase OCT4 expression. The OCT4 expression levels recover as mitosis occurs and the cycle repeats for the cell's descendants consistent with experimental results showing OCT4 resets on re-entry to the G1 phase [273, 274]. This is shown for all cells before BMP4 addition in Figure 5.20.

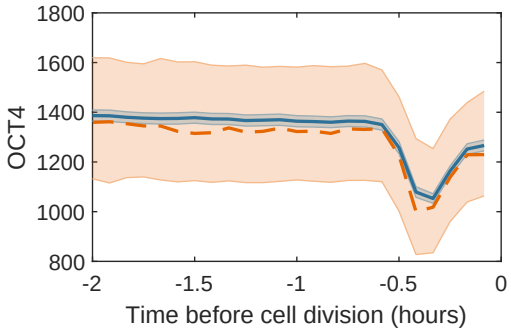

Figure 5.20: The average OCT4 values, mean (blue solid line) with standard error in the mean error bars, and median (orange dashed line) with upper and lower quantile error bars.

5.2.4 The 'unknown' cells

We have quantified the behaviour of OCT4 using a variety of mathematical techniques, some of which characterise the difference between pluripotent and differentiated cells. To illustrate this, we will use the unknown cells (unable to be classified as either pluripotent or differentiated based on their final CDX2 levels) and compare their time series parameters to the pluripotent and differentiated cells.

Firstly, the distribution of all OCT4 values for the unknown cells lies between that for the pluripotent and differentiated cells, both pre- and post-BMP4, as shown in Figure 5.21(a) and (b). This is unsurprising as these cells had an OCT4 expression (along with a CDX2 expression) that did not correspond to either cell fate. Having measured the OCT4 time series for enough time steps to get a distribution of OCT4, comparison could be made to identify whether it best corresponds to that of the pluripotent or the differentiated cell fates, even before any differentiation stimulus is added. However, we still have a large portion of cells which fall in the middle of the two categories (the unknown cells), unable to be confidently classified.

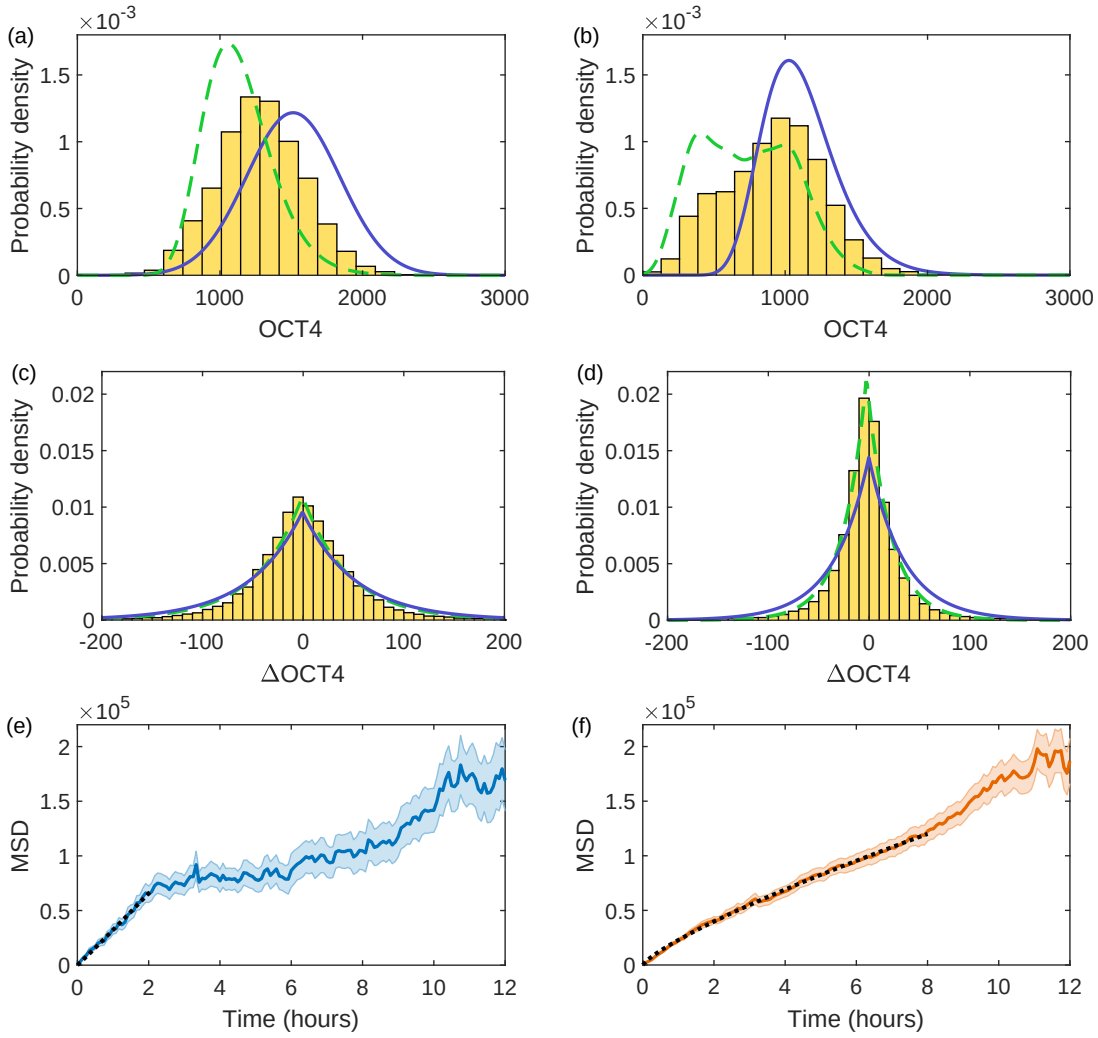

Figure 5.21: The distribution of all OCT4 values for unknown cells (a) pre- and (b) post-BMP4. The corresponding distributions for pluripotent and differentiated cells as a purple solid line and green dashed line, respectively. The distribution of all changes in OCT4, ΔOCT4, for all unknown cells (c) pre- and (d) post-BMP4. The corresponding distributions for pluripotent and differentiated cells as a purple solid line and green dashed line, respectively. The MSD for unknown cells (e) pre and (f) post-BMP4 with error bars the standard error. The black dashed lines show the fits MSD $= \beta t^{\alpha}$ with the parameters (e) $\alpha = 1.03 \pm 0.05$, $\beta = 32000 \pm 1000$ and (f) $\alpha = 0.80 \pm 0.01$, $\beta = 23000 \pm 400$.

There is also a significant difference in the ΔOCT4 distributions post-BMP4, shown in Figure 5.21(d) for the unknown cells. The distribution fits a Laplace distribution with the parameters $\mu^{\dagger} = -1.8$ and $b = 27$, both between their pluripotent and differentiated counterparts. The distributions pre-BMP4 are not different enough to distinguish between the fates, shown in Figure 5.21(c).

There is no distinguishable difference between the cell fates using an autocorrelation analysis or the Hurst exponent (with $\overline{H}$ within errors for all cell fates, including the unknown cells). However, the MSD plots to identify sub-diffusion show a significant difference between the fates, with the MSD for pluripotent cells well described by a power law relationship. The MSD for the unknown cells is shown in Figure 5.21(e) and (f). Pre-BMP4, the MSD is diffusive up to two hours with MSD $= \beta t^{\alpha}$, $\beta = 32000 \pm 1000$ and $\alpha = 1.03 \pm 0.05$, similar to the differentiated cells. Post-BMP4, the relationship fits a power law up to eight hours, with $\beta = 23000 \pm 400$ and $\alpha = 0.80 \pm 0.01$.

These results show that overall, the unknown cells behaviour lies between that of the pluripotent and differentiated cell fates. The unknown pro-fate cells could be the result of a mixture of both populations with the ability to express high and low OCT4 expressions, or cells undergoing a transition phase between pluripotent and differentiated.

5.3 Discussion and conclusions

Promising clinical applications of hESCs require tight control over the pluripotency of hESC colonies. It has been shown that even small PTF fluctuations can bias cell fate decisions and that PTFs are inherited asymmetrically upon cell division [41, 260–262]. It is therefore necessary to quantify the dynamics of key PTFs to further our understanding of how pluripotency is regulated and assist in the development of mathematical modelling. Thorough quantification also provides the basis for experimental comparisons and the identification of systematic and universal behaviours. Here we have used a published data set from [41] to analyse and quantify the dynamics of the PTF OCT4.

The colony considered here grows exponentially, with changing proportions of pluripotent, differentiated and unknown cells. Snapshots of the colony show some spatial patterning of the OCT4 abundance (Figure 5.4), with higher levels of expression of OCT4 visible in cells clustered in the colony centre. A spatial analysis of the colony can be found in [187]. Here we have focused on the quantification of the temporal dynamics.

Time-lapse experiments such as the one considered here [41] provide opportunities for the quantification of temporal PTF regulation which can be compared to, and enhance, current biological knowledge. For example, a sharp decline in OCT4 levels occurring before cell division is noted in [41], in keeping with the transcription factor down regulation known to occur before mitosis [270, 271]. Here we can quantify the phenomena, with the decrease

in OCT4 beginning, on average, 35 minutes (0.58 hours) before cell division, lasting for 15 minutes (0.25 hours), and showing a reduction of 22% from the interphase OCT4 expression. The OCT4 expression levels recover post-mitosis and the cycle repeats for the cell's descendants. This is consistent with experimental results showing OCT4 resets on re-entry to the G1 phase [273, 274].

The study in [41] reveals that sister cells show more closely related OCT4 values than pairs of random cells. Here we take this a step further by quantifying their temporal dynamics in relation to one another. Taking into account any common trends which may affect both cells due to their shared environment, the sister cells before BMP4 show a moderate correlation with each other with a correlation coefficient of 0.5. This is reduced to a slight correlation for pairs that exist after the BMP4 addition (0.3). The fact that these correlations still occur after de-trending further highlights the inherent similarities between related cells. We then consider the OCT4 behaviour over cell lifetimes to explore the manner in which this drift in similarity occurs. The behaviour is summarised in the schematic in Figure 5.22.

Stochastic fluctuations in OCT4 have been shown to bias cell fate [260] with evidence of asymmetric noise leading to noise-mediated cell plasticity [275]. Here we see the change in OCT4 between each 5 minute time interval is isotropic, with an average of zero. A natural assumption in model development would be to simulate this symmetric time-step change in OCT4 with a Normal distribution, however the distribution of all these changes best fits a Laplace distribution. Laplace distributions have previously been applied to gene expression data, with the suggestion that the distribution can represent mixtures of other distributions (e.g., Normal, Pareto) also related to gene expression [276]. The parameters can be used for experimental comparisons, as direct inputs into computational models and for model verification. Further experimental data (e.g., different cell lines, culture conditions, restricted geometries, cell densities) are needed confirm the robustness, estimate the parameters for other experimental conditions and investigate how this is affected by cell-cell interactions. This only allows us to capture the nature of the variation in OCT4 and further aspects of the behaviour need to be considered to fully describe the OCT4 regulation over time.

Although this shows that overall, positive changes in OCT4 are just as likely to occur as negative ones, it does not reveal anything about the temporal nature of these fluctuations and hence any correlation properties which may be evident over time (for example, all the positive changes in OCT4 could come one after the other, followed by all the negative changes, it does not mean that a positive change is necessarily followed by a negative change). There is also a difference in these fluctuations after the differentiation agent, with the addition of BMP4 provoking tighter self-regulation across all cell fates. Experiments using increasing numbers of cells and colony sizes (using varying size multi-well growth

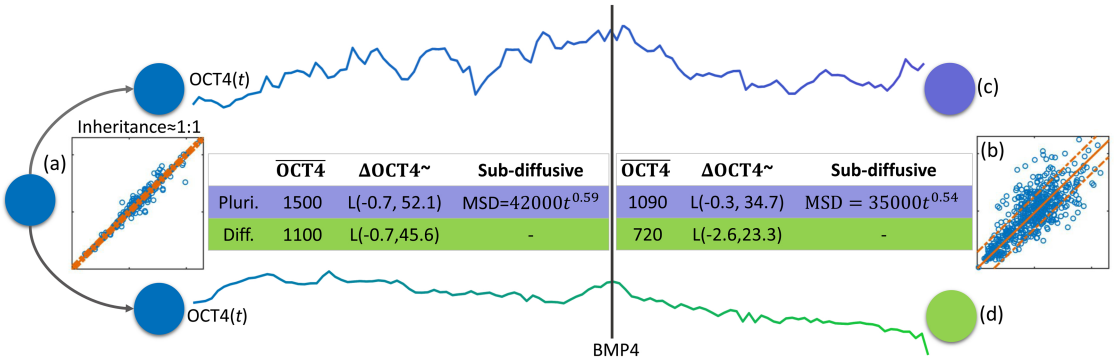

Figure 5.22: An illustration of the dynamics in OCT4 over a cell lifetime. (a) OCT4 is split, possibly asymmetrically but on average in a 1:1 ratio [41] before fluctuating in a sub-diffusive manner (with a power law relationship for pluripotent cells), resulting in (b) more variation in sister cells at the end of their lifetimes (over an average of 14 hours [41]). This can result in cells of different fates: (c) pluripotent and (d) differentiated.

plates and boundaries) can be used to investigate whether this self-regulation is a collective behaviour effect that begins at a certain colony size.

The distributions of OCT4 provide a quantitative comparison for future experiments, characterising the spread and skew of OCT4 expression. Similarly, the Laplace distribution fittings to the change in OCT4 between the five minute time intervals can be a direct input into computation models requiring ΔOCT4 as a probabilistic parameter. The distributions can also be used for model verification. Regardless of the modelling method used, the distributions of OCT4 and ΔOCT4 produced via a simulation can be compared, both qualitatively, and quantitatively by the parameters of the distribution, to these experimental results.

A significant finding of this analysis is the quantification of the self-regulatory properties of OCT4 within cells. We use the Hurst exponent, previously used to characterise gene expression [277], DNA sequencing [278], stem cell division times [279] and self-renewal capacities [280]. The Hurst exponent quantifies the level of anti-persistence in the OCT4 regulation (0.38). Broadly, the identification of a Hurst exponent which is not Brownian suggests the use of specific equations, i.e., the equation needs to be able to produce this anti-persistence. Furthermore, it can be a direct input into some stochastic modelling equations in which the parameter, H, is required, e.g., fractional Brownian motion [103, 266, 268, 269]. We also use Poincaré plots to visualise the OCT4 values in relation to the previous snapshot but note that other methods are also applicable here, such as diffusion maps [281].

An autocorrelation analysis shows significant anti-correlation, in keeping with the regulation of PTFs [43, 156, 242]. Throughout the colony growth, anti-correlation of at least five hours is seen in 86% of cells (with no significant difference between the cell

fates, suggesting OCT4 regulation is quantitatively comparable in pluripotent and differentiated cells), and on average occurs between 3 and 12 hours into a cell's lifetime, with cells in an anti-correlated state for 60% of their cell cycle. The results suggest that the anti-correlation is an inherent property of the cells, across all cell fates. Further experiments are needed to clarify that this is the case under different experimental conditions (i.e., different sized colonies, other cell lines, in different geometries) but this provides a further quantitative statistic for comparisons. The identification of this systematic property has implications for the underlying stochastic chemistry of the OCT4 regulation and can be used to inform chemical models of PTF regulation, often based on the Hill equations [147, 148].

This is further illustrated by considering the behaviour of the cells in the diffusion framework, with cells across all fates showing significant sub-diffusivity. For pluripotent cells, the sub-diffusivity can be described by a power-law relationship. The sub-diffusive nature of the time series allows the complete characterisation using only two summary parameters, α and β. This highlights the presence of a universal behaviour in the cells, and can be used as a direct parameter input into Brownian (and the fractional and geometric extensions) computational models allowing the future behaviour of the OCT4 regulation to be predicted statistically.

It is worth noting that the anti-persistence and sub-diffusivity is not contradictory to the conclusion of pre-determined cell fate in [41], as here we consider the behaviour over individual cell lifetimes, (i.e., shorter time scales) and do not take into account other behaviours in the colony (i.e., over multiple cell divisions, longer time scales).

The experiment in [41] has led to a rich analysis, allowing us to establish the language through which to guide mathematical modelling; informing equation choices, revealing properties of the inherent chemistry and identifying universal behaviours. In general, this highlights the need for further temporal experimental data on OCT4 and other transcription factors. In Chapter 6, we will use these results to develop mathematical models to describe temporal OCT4 regulation.

Chapter 6

Modelling pluripotency

An introduction to mathematical models of the pluripotency regulatory network is given in Chapter 1. Here we simplify the process of modelling pluripotency by focussing on the single pluripotency transcription factor, OCT4. We use the quantitative analysis from Chapter 5 and Experiment 3 (details given in Appendix A) to explore different mathematical models for its temporal regulation and characteristic fluctuations. The OCT4 expression for pluripotent and differentiated cells in Experiment 3 is shown in Figure 6.1. Note that at 40 hours the differentiation agent BMP4 was added to the cells to facilitate differentiation, resulting in a decline in OCT4 expression in some cells.

Since pluripotency (and hence OCT4) is inherently stochastic [41, 60, 240, 275], we focus on different forms of stochastic models to capture the behaviour, namely: fractional Brownian motion and the stochastic logistic equation. We consider different methods of noise production (additive and multiplicative) and their effects on temporal OCT4 simulation. We also examine the use of shifting carrying capacities and Allee effects to simulate a reduction in pluripotency towards the differentiated state.

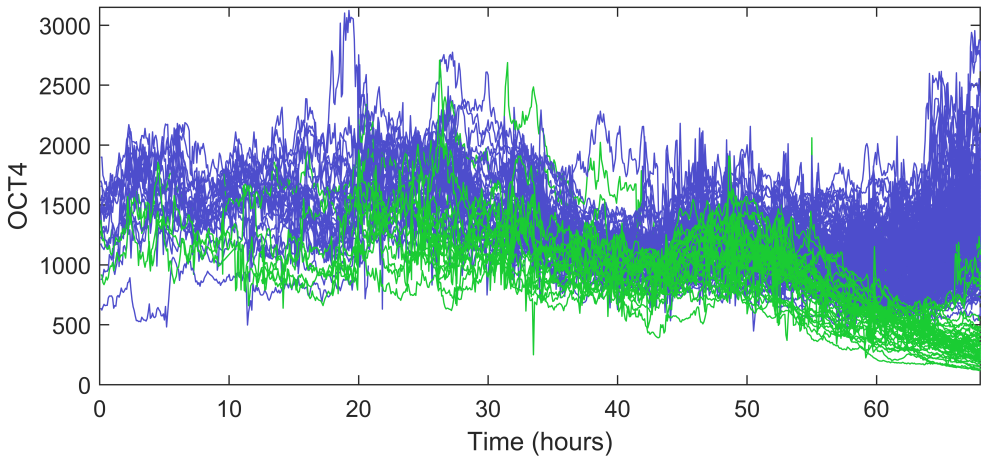

Figure 6.1: The OCT4 expression for all pluripotent (purple) and differentiated (green) cells from Experiment 3 (details given in Chapter 5 and Appendix A). The differentiation agent BMP4 was added at 40 hours causing a reduction in OCT4 in the differentiated cells.

6.1 Modelling temporal OCT4

6.1.1 Basic model

In the following sections we explore the use of different stochastic models as a framework for temporal OCT4 regulation, aiming to capture the experimental behaviour shown in Figure 6.1. All the models discussed have the same basis, with the initial conditions and cellular division incorporated using the algorithmic base model below.

1. We begin with a chosen initial number of cells, $N = N_0$.
2. Each of the N cells are allocated an initial OCT4 value. This is extracted probabilistically from the kernel density fitting to the distribution of OCT4($T = 0$) from Experiment 3, shown in Figure 6.2(a).
3. Each of the N cells are allocated a cell cycle duration. This is extracted probabilistically from the kernel density fitting to the distribution of cell cycle times for all pre-BMP4 cells from Experiment 3, shown in Figure 6.2(b). Each cell's position in its cell cycle is chosen uniformly.
4. For each of the N cells the OCT4 values for the duration of their cell cycle are simulated using one of the stochastic models.
5. Each of the N cells divide into two cells at the end of their cell cycle. For each of the two daughter cells, their initial OCT4 value is set to the final OCT4 value of the mother cell.
6. Steps 4 and 5 are repeated for the number of required division events.

When the cell cycle times are generated in step 3 it is necessary to specify how much of the cell cycle has already elapsed. If all cells begin at the start of their cell cycle at the start of the simulation then divisions will be synchronised, visible as 'steps' in the number of cells over time, $N(t)$, as shown in Figure 6.3(a). Avoiding this synchronisation by starting cells at different points in their cell cycle gives a more accurate representation of colony size, as shown in Figure 6.3(b).

Although we have used the analysis of Experiment 3 to inform the initial conditions and the cell cycle simulation, this is flexible and can easily be adapted to other experimental results. The OCT4 regulation itself is captured in step 4 and is open to many mathematical modelling techniques. In the next section we consider using fractional Brownian motion and the stochastic logistic equation for OCT4 dynamics pre-BMP4 addition.

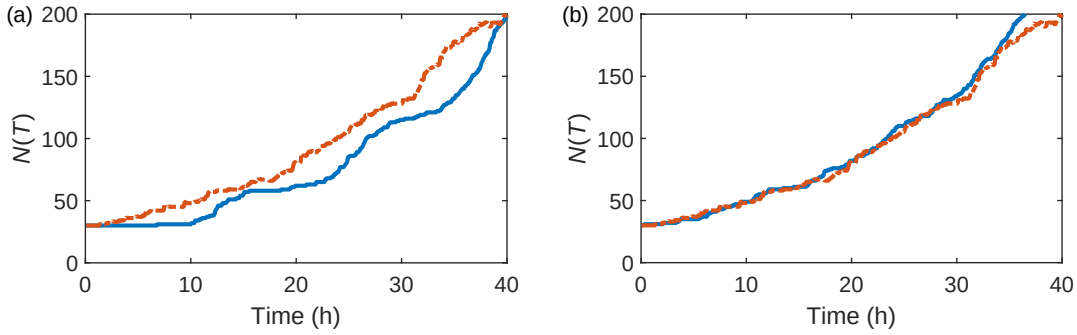

Figure 6.3: The number of cells over time, $N(t)$, when (a) cellular division is synchronised and (b) cellular division is not synchronised in step 3 of the common base model. Blue solid lines show the simulated population sizes and orange dashed lines the population from Experiment 3.

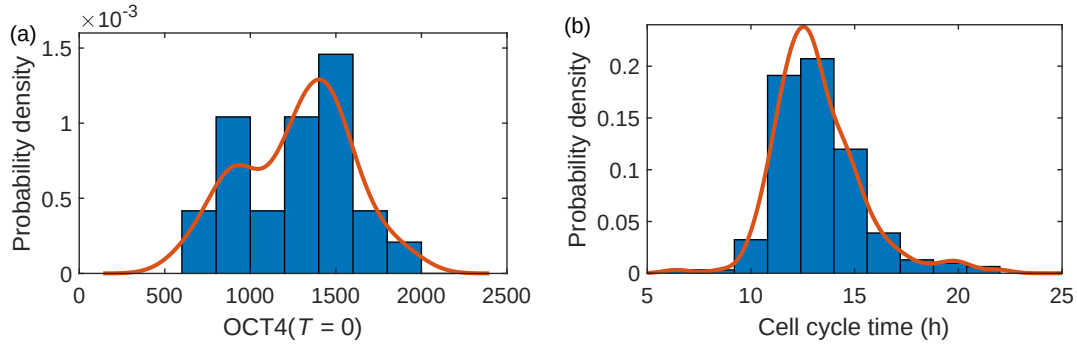

Figure 6.2: The distributions used to inform the initial conditions and cellular division in the common base model. (a) The distribution of initial OCT4 values, OCT4($T = 0$), from Experiment 3 with kernel density fitting shown in orange. (b) The distribution of cell cycle duration times for all cells before BMP4 addition from Experiment 3, with kernel density fitting shown in orange.

6.1.2 Fractional Brownian motion

The analysis of Experiment 3 presented in Chapter 5 revealed that temporal OCT4 is anti-persistent with an average Hurst exponent of $H = 0.38$. The Hurst exponent $H \neq 0.5$ indicates that the fluctuations in OCT4 cannot be captured by simple Brownian motion. Instead we can consider the generalisation, fractional Brownian motion (fBm). Unlike Brownian motion, fBm allows for non-independent increments and hence persistence or anti-persistence. The random function for fBm over time t, $B_H(t)$, with an initial value

$B_H(0)$ is defined by

$$B_H(t) = B_H(0) + \frac{1}{\Gamma(H+0.5)} \int_{-\infty}^{0} \left[(t-s)^{H-0.5} - (-s)^{H-0.5}\right] dB(s) + \frac{1}{\Gamma(H+0.5)} \int_{0}^{t} (t-s)^{H-0.5} dB(s),$$

where H is the Hurst exponent and Γ is the gamma function. There are several different ways to simulate fBm, both exact and approximate [282–284]. Here we use the Matlab function *ffgn* [285] which uses the circulant embedding technique for $H < 0.5$ [286] and Lowen's method [287] for $H > 0.5$ (both exact methods) to simulate the fractional Brownian noise. There is also an inbuilt Matlab function *wfbm* (available in the Wavelet toolbox) which uses a wavelet based approximate simulation method [288].

Example realisations of the fractional noise for fBm processes with $H = 0.1, 0.5$ and 0.9 are shown in Figure 6.4(a-c). The corresponding fBm functions with the initial condition $B_H(0) = 0$ are shown in Figure 6.4(d-f). A low Hurst exponent (anti-persistence) results in a reversion to the mean, whilst a high Hurst exponent (persistence) results in periods of continued trends. We can use fBm to simulate OCT4 over time (step 4 of the base model) with a scaling parameter σ which controls the level of noise, i.e., σB_H. The simulations shown in Figure 6.4(a-f) use $\sigma = 1$. Each time series for OCT4 can then be generated as $\text{OCT4}(0) + \sigma B_H$. Realisations of OCT4 over 40 hours for ten initial cells with $\sigma = 90$ (extracted from Experiment 3 as the standard deviation of ΔOCT4 for all pre-BMP4 cells), OCT4(0) sampled from the experimental distribution shown in Figure 6.2(a), and $H = 0.1, 0.5$ and 0.9 are shown in Figure 6.4(g-i) to illustrate the effect of the Hurst exponent.

We extracted a mean Hurst exponent of 0.38 for all cells in Experiment 3 (given in Table 5.6). This experimental OCT4 expression up to 40 hours is shown with time in Figure 6.5(a) and as a distribution in Figure 6.5(b). A comparable simulation using fBm with 16 initial cells, $H = 0.38$, and $\sigma = 90$ is shown in Figure 6.5(b) and (d). It is clear that the regulation from anti-persistence is not sufficient to keep the OCT4 expression within the range seen in the experiment.

One possible method of limiting the range of OCT4 is to impose boundary conditions. We have two choices for boundary conditions: absorbing and reflecting. Absorbing boundary conditions imply that once the OCT4 level reaches the boundary, the cell is somehow removed from the experiment. This could biologically correspond to cell death, but we don't see any indication of particularly high or low OCT4 expressions resulting in cell death in Experiment 3. The OCT4 simulation for fBm with 16 initial cells, $H = 0.38$, $\sigma = 90$, and absorbing boundary conditions at zero and 2500 is shown in Figure 6.5(e) and (f).

Reflecting boundary conditions imply that when the OCT4 expression reaches the boundary, it is reflected back in the opposite direction. Biologically this corresponds to an additional regulatory effect; if the OCT4 level in a cell becomes too low, there is systematic regulation to increase it (and vice versa). The simulation using fBm with 16 initial cells, $H = 0.38$, $\sigma = 90$, and reflecting boundary conditions at zero and 2500 is shown in Figure 6.5(g) and (h). Reflecting boundary conditions produce a result more similar to the experiment than absorbing boundary conditions since cells are not artificially removed, but it still creates a harsher boundary than is seen in the experiment. Additionally, although the boundary conditions somewhat artificially force the OCT4 into the desired range, the spread of the overall expressions is not well captured, shown in the corresponding histograms in Figure 6.5(b, f) and (h).

This illustrates that the anti-persistence from the Hurst exponent alone in fBM is not enough to capture the OCT4 regulation seen in the experiment, even with boundary conditions. The imposition of any boundary conditions also requires further investigation to elucidate their nature, positioning and the biological implications. However, we can still incorporate fBm noise into other models to generate the anti-persistence. In the next section we consider describing temporal OCT4 with the stochastic logistic equation and explore the regulatory efforts of a limiting carrying capacity.

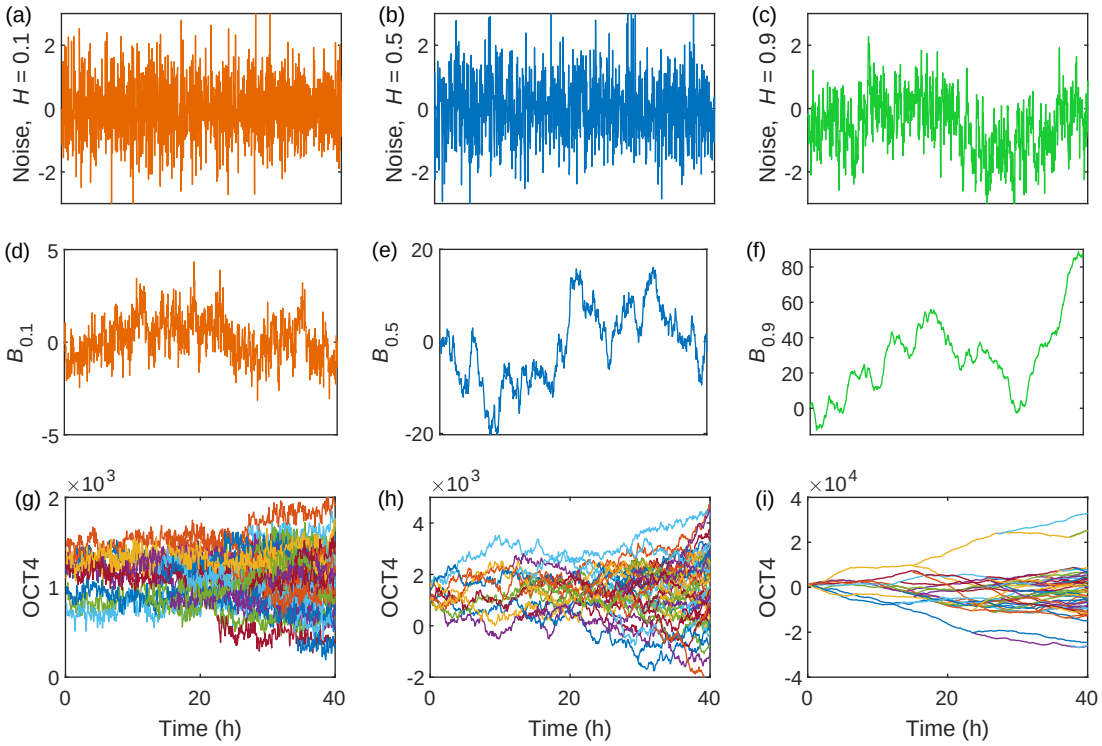

Figure 6.4: Realisations of simulated noise in Fractional Brownian motion with (a) $H = 0.1$ (anti-persistence), (b) $H = 0.5$ (Brownian) and (c) $H = 0.9$ (persistence), (d-f) the corresponding simulated trajectories with initial condition $B_H(0) = 0$. (g-i) Simulation of OCT4 for 40 hours, with ten initial cells, and temporal OCT4 determined by simulated realisations of σB_H with $\sigma = 90$ and (g) $H = 0.1$, (h) $H = 0.5$ and (i) $H = 0.9$.

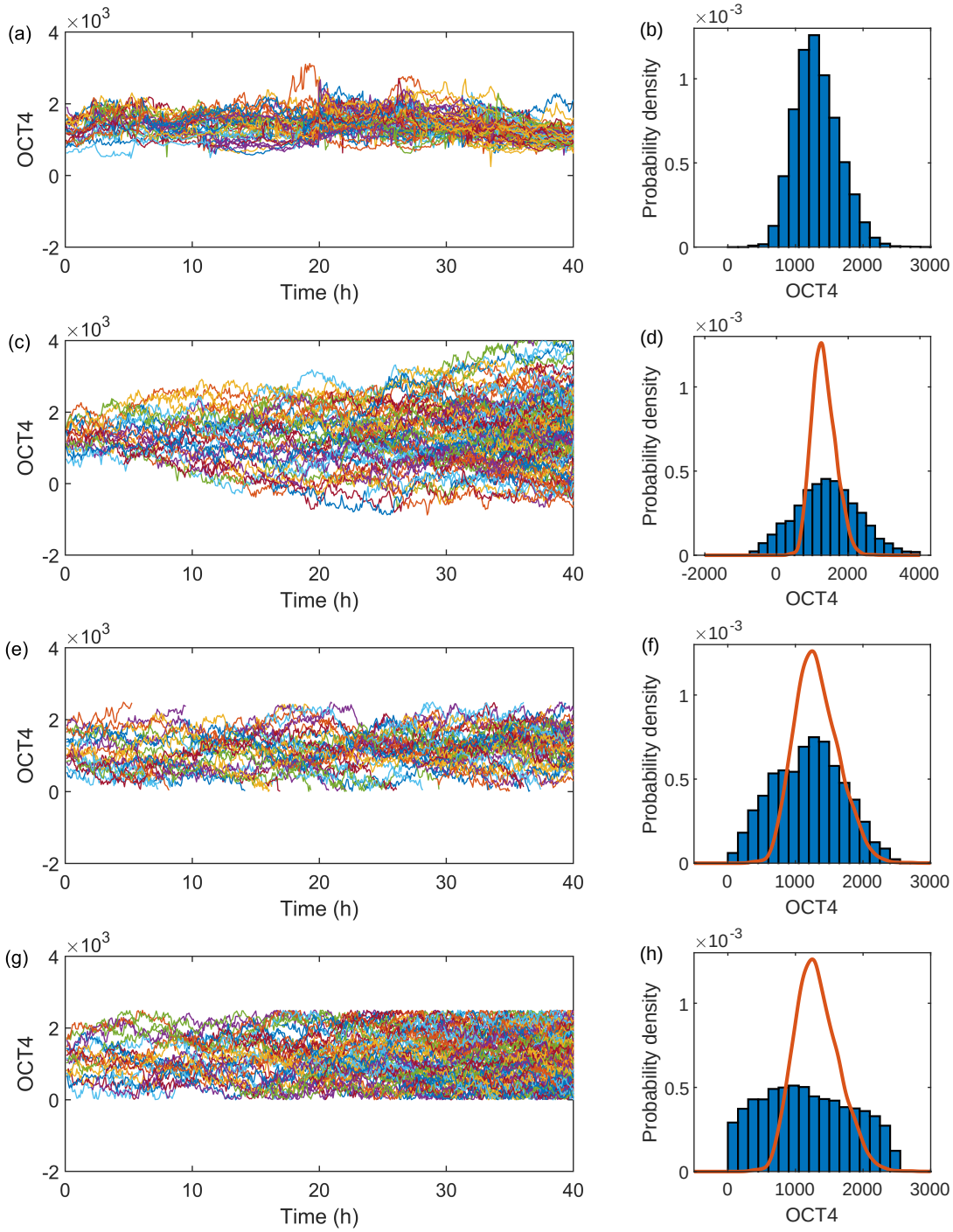

Figure 6.5: (a) Experimental OCT4 expression for pluripotent and differentiated cells between zero and 40 hours. Colours correspond to different cells. Simulated OCT4 expression using fBm with 16 initial cells, $\sigma = 90$ and $H = 0.38$ with (c) no, (e) absorbing, and (d) reflecting boundary conditions at zero and 2500. The corresponding histograms for (b) experimental OCT4 and simulated OCT4 with (d) no, (f) absorbing, and (h) reflecting boundaries. The kernel density fitting to the experimental distribution is shown in orange.

6.1.3 The stochastic logistic equation

We introduced the stochastic logistic equation (SLE) and its use for modelling cell populations in Chapter 1. In this section we explore its novel application to simulating temporal OCT4 regulation. We previously saw that fBm alone does not fully capture the regulatory behaviour of OCT4 seen in Experiment 3. This suggests the use of some additional regulatory method, such as a carrying capacity term as included in the SLE. We consider the SLE with additive noise, multiplicative noise, and the effect of a time-dependent carrying capacity.

The SLE with additive noise

There are several ways stochasticity can be introduced into the logistic equation, e.g., additive noise, multiplicative noise, a noisy growth rate parameter r or carrying capacity K. The most straightforward of these is additive noise which can be introduced by adding a noise term to the differential equation. The SLE with additive noise to describe OCT4, O, over time, t, is then

$$\frac{dO}{dt} = rO\left(1 - \frac{O}{K}\right) + \sigma_{\mathrm{A}}\xi, \tag{6.1}$$

where ξ is the stochastic noise (e.g., Wiener/Brownian noise, or fBM noise) and σ_{A} is a scaling parameter controlling the magnitude of the noise.

We can estimate the parameters for Eq (6.1) from Experiment 3 shown in Figure 6.6(a) and (b). The noise ξ corresponds to fBm noise with a Hurst exponent $H = 0.38$ and the scaling parameter can be estimated as the standard deviation of ΔOCT4, $\sigma_{\mathrm{A}} = 90$. We can also estimate the carrying capacity as the median of all the experimental OCT4 values, $K = 1290$, shown with the experimental distribution in Figure 6.6(b).

The OCT4 dynamics simulated using Eq (6.1) with $r = 0.02$ are shown in Figure 6.6(c) and (b). Although the regulatory effect of the carrying capacity works well to capture the upper bound of OCT4 expression, an additional boundary condition at zero is still required (if the stochasticity gives rise to $O < 0$ then $dO/dt < 0$ resulting in $O \to -\infty$). A distinguishing feature not captured by the model is the positive skew in the distribution of all occurring OCT4 values, visible in Figure 6.6(b) and (d). The model promotes tighter regulation above the carrying capacity than below it, resulting in fewer OCT4 expressions above the carrying capacity than seen experimentally. This suggests that the stochasticity has some dependence on the state of the system, which could be introduced using multiplicative noise.

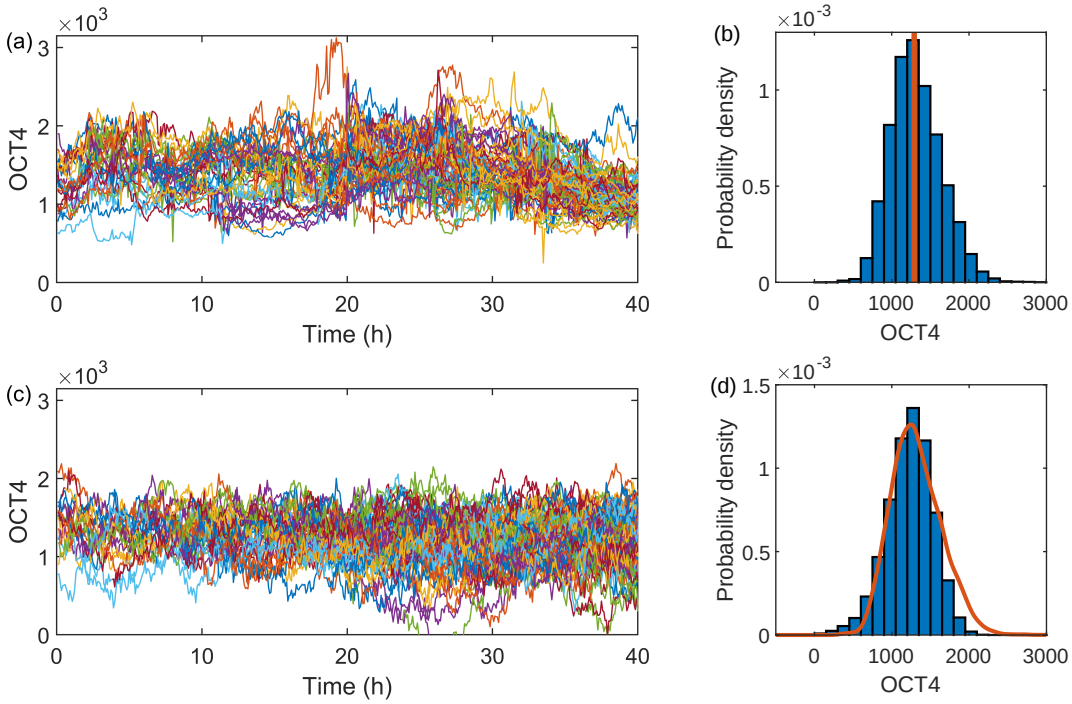

Figure 6.6: (a) Experimental OCT4 expression for pluripotent and differentiated cells between 0 and 40 hours. Colours correspond to different cells. (b) The corresponding histogram of experimental OCT4 expression with the orange line showing the median (and estimated carrying capacity) of 1290. (c) Simulated OCT4 expression using the SLE, Eq (6.1), with 16 initial cells, $r = 0.02$, $K = 1290$, $\sigma_A = 90$ and $H = 0.38$, with an absorbing boundary condition at zero. (d) The corresponding histogram of simulated OCT4 expression using Eq (6.1) with the experimental distribution in orange.

The SLE with multiplicative noise

Another method of adding stochasticity to the logistic equation is by introducing multiplicative noise. Whereas the additive noise in Eq (6.1) has no dependence on the state of the system and corresponds to making dO/dt symmetrically noisy, multiplicative noise changes depending on the current conditions. In the case of our temporal OCT4 simulation, multiplicative noise can be used to generate noise which has a greater magnitude when the system is close to the carrying capacity (thus resulting in more stochastically high OCT4 expressions) and a reduced magnitude when far away from the carrying capacity. Hints of this behaviour can be seen in Figure 6.6(a), with larger fluctuations apparent in the cells exhibiting above average OCT4 expression.

For simulating the SLE with multiplicative noise we first consider the rearrangement of the logistic equation,

$$\frac{1}{O}\frac{dO}{dt} = r\left(1 - \frac{O}{K}\right), \tag{6.2}$$

which can be written as

$$\frac{d\log(O)}{dt} = r\left(1 - \frac{O}{K}\right). \tag{6.3}$$

Applying the substitution $X = \log(O)$ and adding stochasticity ξ with noise scaling parameter σ_{M} gives

$$\frac{dX}{dt} = r\left(1 - \frac{e^X}{K}\right) + \sigma_{\mathrm{M}}\xi, \tag{6.4}$$

which can then be used to simulate $X = \log(O)$, with the dynamics of OCT4 recovered from $O = e^X$. Example realisations of Eq (6.4) for both X and O are shown in Figure 6.7(a) and (b), respectively, to illustrate the effect of multiplicative noise in a typical logistic growth scenario for varying σ_{M}. The result is amplified noise for stochasticity occurring above the carrying capacity.

The temporal OCT4 dynamics simulated using the SLE with multiplicative noise, Eq (6.4), with ξ fBM noise with $H = 0.38$, $r = 0.005$, $K = 1290$ and $\sigma_{\mathrm{M}} = 0.0045$ for 16 initial cells are shown in Figure 6.8(c). The multiplicative noise results in cells with expressions above the carrying capacity exhibiting increased stochasticity, with lower expression cells showing tighter regulation. The simulated distribution has a slight positive skew, and is qualitatively similar to the experimental distribution as shown in Figure 6.8(d).

This model provides a good basis for capturing the experimental results across the whole time period and is an improvement on the SLE with additive noise. However, we should also examine the model's performance at various instances in time. The positive skew in the distribution of OCT4 expression arises from the pluripotent cell group, the distributions for which are shown for all times in Figure 5.8(a), and for times between zero and 40 hours in Figure 6.9(a). The distributions of OCT4 for differentiated cells are approximately symmetric across all times as shown in Figure 5.8(c) and Figure 6.9(c). The corresponding temporal distributions from Eq (6.4) show consistent skew across all times as shown in Figure 6.9(c).

This indicates that firstly, the model should consider the cell fates separately as only the pluripotent cells show significant skewness and secondly, the model needs to incorporate a method of shifting this skew temporally for the pluripotent cells. In the following sections we discuss two methods of doing this: the SLE with a transition between dominant additive and dominant multiplicative noise, and the SLE with a time dependent carrying capacity.

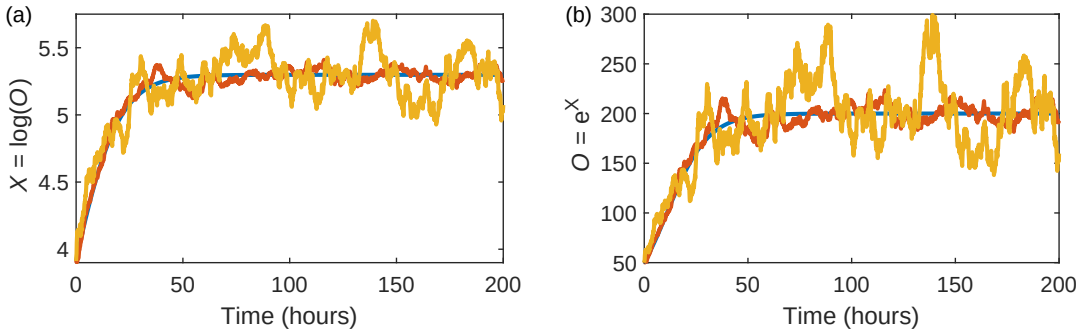

Figure 6.7: The SLE with multiplicative noise. Realisations of the dynamics of (a) $X = \log(O)$ and (b) $O = e^X$ from Eq (6.4) with $r = 0.1$/h, $K = 100$, $\xi = W_t$ and $\sigma_{\mathrm{M}} = 0$ (blue), 0.025 (orange) and 0.075 (yellow).

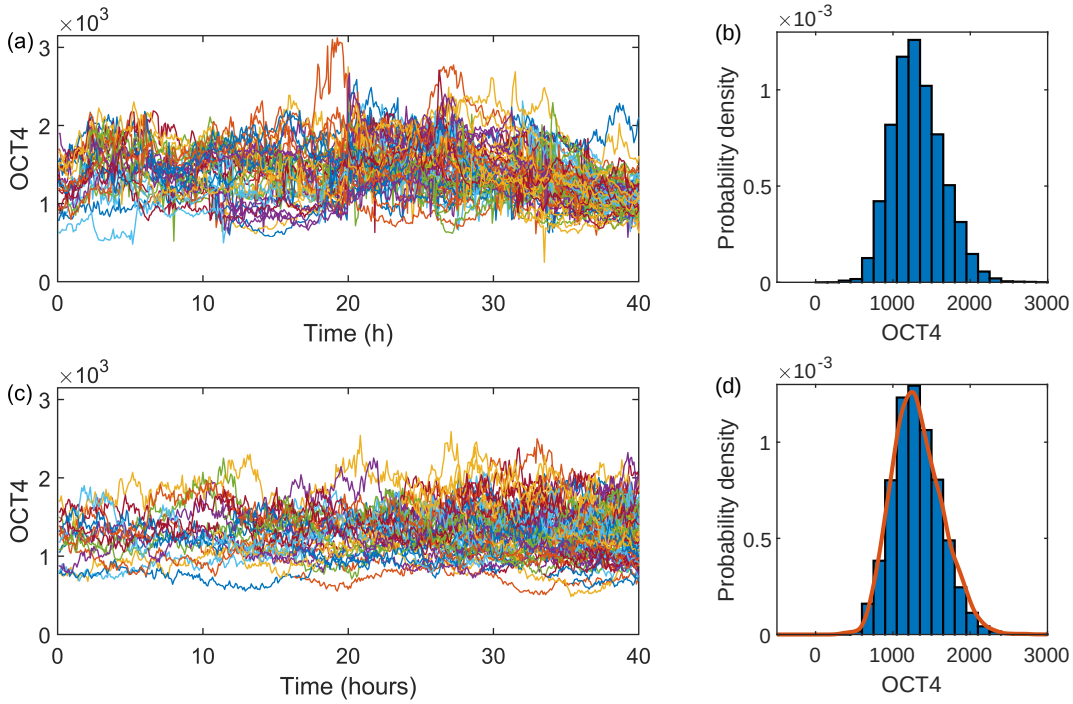

Figure 6.8: (a) Experimental OCT4 expression for pluripotent and differentiated cells between 0 and 40 hours. Colours correspond to different cells. (b) The corresponding histogram of experimental OCT4 expression. (c) Simulated OCT4 expression using the SLE with multiplicative noise, Eq (6.4), with 16 initial cells, $r = 0.005$, $K = 1290$, $\sigma_{\mathrm{M}} = 0.0045$ and ξ fBM noise with $H = 0.38$. (d) The corresponding histogram of simulated OCT4 expression with the experimental distribution in orange.

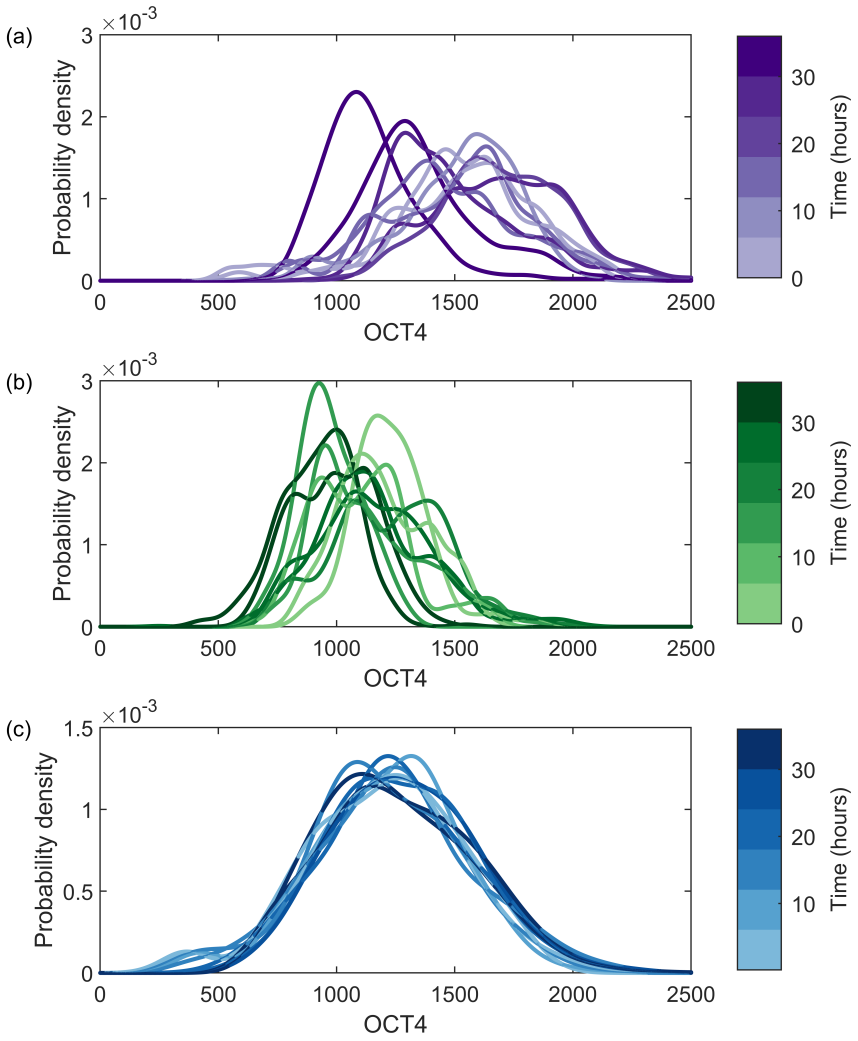

Figure 6.9: The distribution of OCT4 expression for time intervals between zero and 40 hours for (a) pluripotent and (b) differentiated cells from Experiment 3 and (c) simulated cells using the SLE with multiplicative noise, Eq (6.4), with 16 initial cells, $r = 0.005$, $K = 1290$, $\sigma_{\mathrm{M}} = 0.0045$ and ξ fBM noise with $H = 0.38$. from Eq (6.4).

The SLE with additive and multiplicative noise

To capture the changing temporal skew for pluripotent cells, we could include both additive and multiplicative noise. If additive noise is dominant at early times, and multiplicative noise at later times, the resulting OCT4 distribution will be symmetric at early times and skewed at later times. We can consider the following rearrangement of the stochastic logistic equation with additive noise

$$\frac{d\log(O)}{dt} = r\left(1 - \frac{O}{K}\right) + \frac{\sigma_{\mathrm{A}}}{O}\xi_1, \tag{6.5}$$

make the substitution $X = \log(O)$ and introduce the multiplicative noise term $\sigma_M \xi_2$,

$$\frac{dX}{dt} = r\left(1 - \frac{e^X}{K}\right) + \frac{\sigma_A}{e^X}\xi_1 + \sigma_M \xi_2. \tag{6.6}$$

As before, we can simulate the dynamics for X and recover the dynamics for $O = e^X$.

For simplicity, we can consider the change between additive and multiplicative noise as a switch: for $0 < t < 20\,\text{h}$, $\sigma_A = 90$ and $\sigma_M = 0$, and for $t > 20\,\text{h}$, $\sigma_A = 0$ and $\sigma_M = 0.03$. The results for the OCT4 dynamics within this regime (with $r = 0.01$, $K = 1290$ and ξ_i fBm noise with $H = 0.38$) are shown in Figure 6.10. The temporal distributions shown in Figure 6.10(c) illustrate the effect of the noise switch with the appearance of a positive skew at later times.

Although this well captures the overall distribution, shown in Figure 6.10(c), and provides the desired temporal change in skew (which could be further smoothed with a more sophisticated time dependent noise function), it does not result in a shift in the mode as drastic as the one apparent in Figure 6.9(a). Another possible method of capturing this behaviour is to make the carrying capacity time dependent.

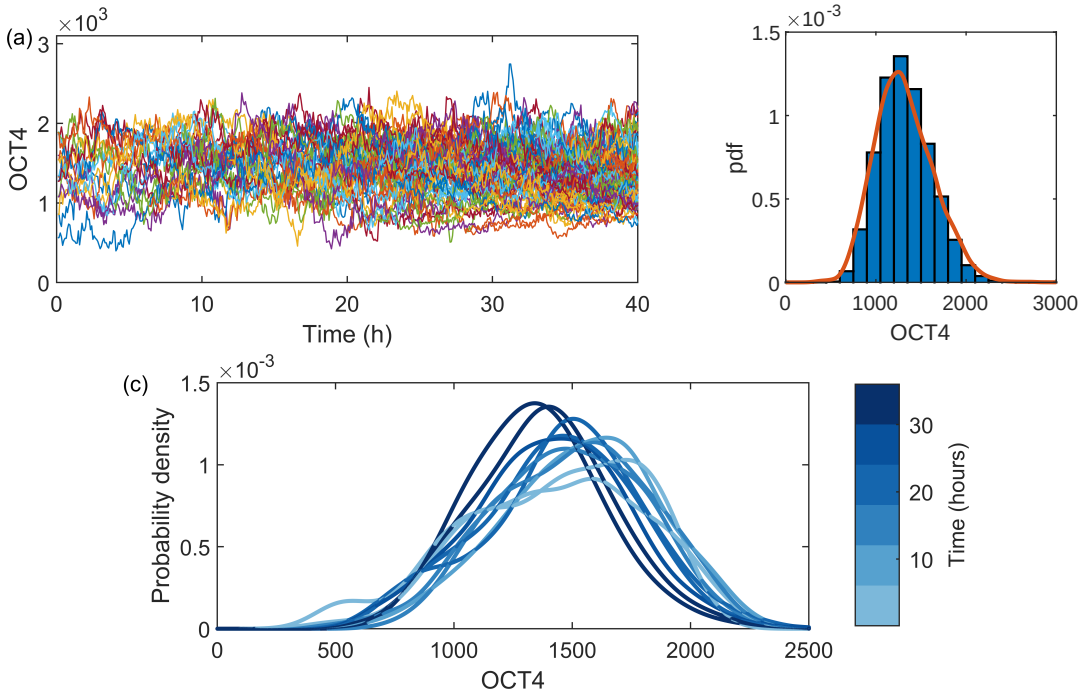

Figure 6.10: The OCT4 dynamics between zero and 40 hours for 16 initial cells following Eq (6.6) with $\sigma_A = 90$ and $\sigma_M = 0$ for $0 < t < 20\,\text{h}$, and $\sigma_A = 0$ and $\sigma_M = 0.03$ for $t > 20\,\text{h}$, $r = 0.01$, $K = 1290$ and $H = 0.38$. (a) OCT4 over time, (b) the corresponding histogram of all simulated OCT4 values (blue) with the experimental distribution (orange) and (c) distributions of the simulated OCT4 split by time intervals.

The SLE with a time-dependent carrying capacity

The distributions of OCT4 at different time intervals reveal that the skew occurs in the pluripotent cells at later times as shown in Figure 6.9. We can further examine this by considering the mean, median, mode and skewness (defined as $E(x-\mu)^2/\sigma^3$ where $E[\cdot]$ is the expected value for variable x with mean μ and standard deviation σ) of the time-discretised OCT4 for both pluripotent and differentiated cells, shown in Figure 6.11. A decrease in the mean, mode and median, and an increase in skewness is seen in the pluripotent cells from around 25 hours onwards. This suggests that the cell fates must be considered separately and that a time-dependent carrying capacity may be appropriate to capture this shift in the averages and mode for the pluripotent cells. We use the stochastic logistic equation for both cell fates, with both multiplicative and additive noise, as in Eq (6.6), and a carrying capacity which varies with time,

$$\frac{dX}{dt} = r\left(1 - \frac{e^X}{K(t)}\right) + \frac{\sigma_{\mathrm{A}}}{e^X}\xi_1 + \sigma_{\mathrm{M}}\xi_2. \tag{6.7}$$

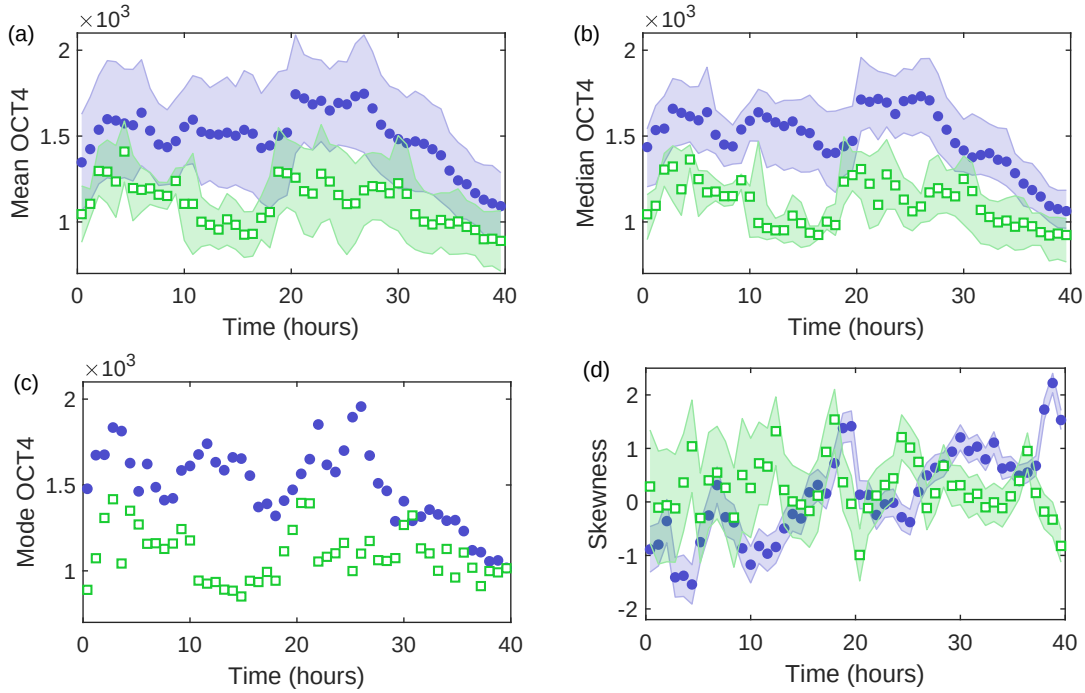

Figure 6.11: For the time-discretized OCT4 expressions from Experiment 3 the (a) mean with standard deviation error bars, (b) median with interquartile range error bars, (c) mode and (d) skewness with standard error bars for pluripotent (purple, filled circles) and differentiated (green, open squares) cells.

We can estimate the carrying capacity as the median OCT4 between zero and 25 hours resulting in $K_\mathrm{p} \approx 1500$ and $K_\mathrm{d} \approx 1100$ for pluripotent and differentiated cells, respectively. For simplicity, post-25 hours, we will estimate both carrying capacities as $K \equiv K_\mathrm{p} = K_\mathrm{d} \approx 1000$. This reduction in the carrying capacity will initiate the corresponding reduction in the mean, median and mode quantities we see experimentally. The OCT4 dynamics using these time dependent carrying capacities in Eq (6.7) for 14 pluripotent and two differentiated cells (corresponding to the experimental N_0) are shown in Figure 6.12 and Figure 6.13. The model parameters for pluripotent and differentiated cells, pre- and post-25 hours are summarised in Table 6.1.

The lower carrying capacity results in consistently lower OCT4 expression for the differentiated cells, as shown in Figure 6.12(a) for the experimental cells and in Figure 6.12(e) for the simulated cells. The reduction in the carrying capacity for both cell fates captures well the resulting reductions in the average OCT4 expressions at later times, as shown in Figure 6.12(b-h). The overall distribution of OCT4 expressions is also well described, as shown in Figure 6.13(a) and (b). The average Hurst exponent is recovered by the inclusion of the fBm noise, with a mean of $\overline{H} = 0.38$ for both experimental and simulated cells, shown in Figure 6.13(c) and (d), respectively. The standard deviation of H for the experimental cells (0.09) is higher than for the simulated cells (0.06). This is not surprising since we only input the average H value to simulate the fBM noise, but experimentally this will be affected by a wide variety of environmental and internal factors.

	Parameter	$t < 25\,\mathrm{h}$	$t \geq 25\,\mathrm{h}$
Pluripotent	N_0	14	
	r	0.015	
	K	1500	1000
	σ_A	30	
	σ_M	0.035	
	H	0.38	
Differentiated	N_0	2	
	r	0.015	
	K	1100	1000
	σ_A	20	
	σ_M	0.03	
	H	0.38	

Table 6.1: Simulation parameters for generating OCT4 expression for pluripotent and differentiated cells using the SLE with additive and multiplicative noise, and a time dependent carrying capacity, Eq (6.7).

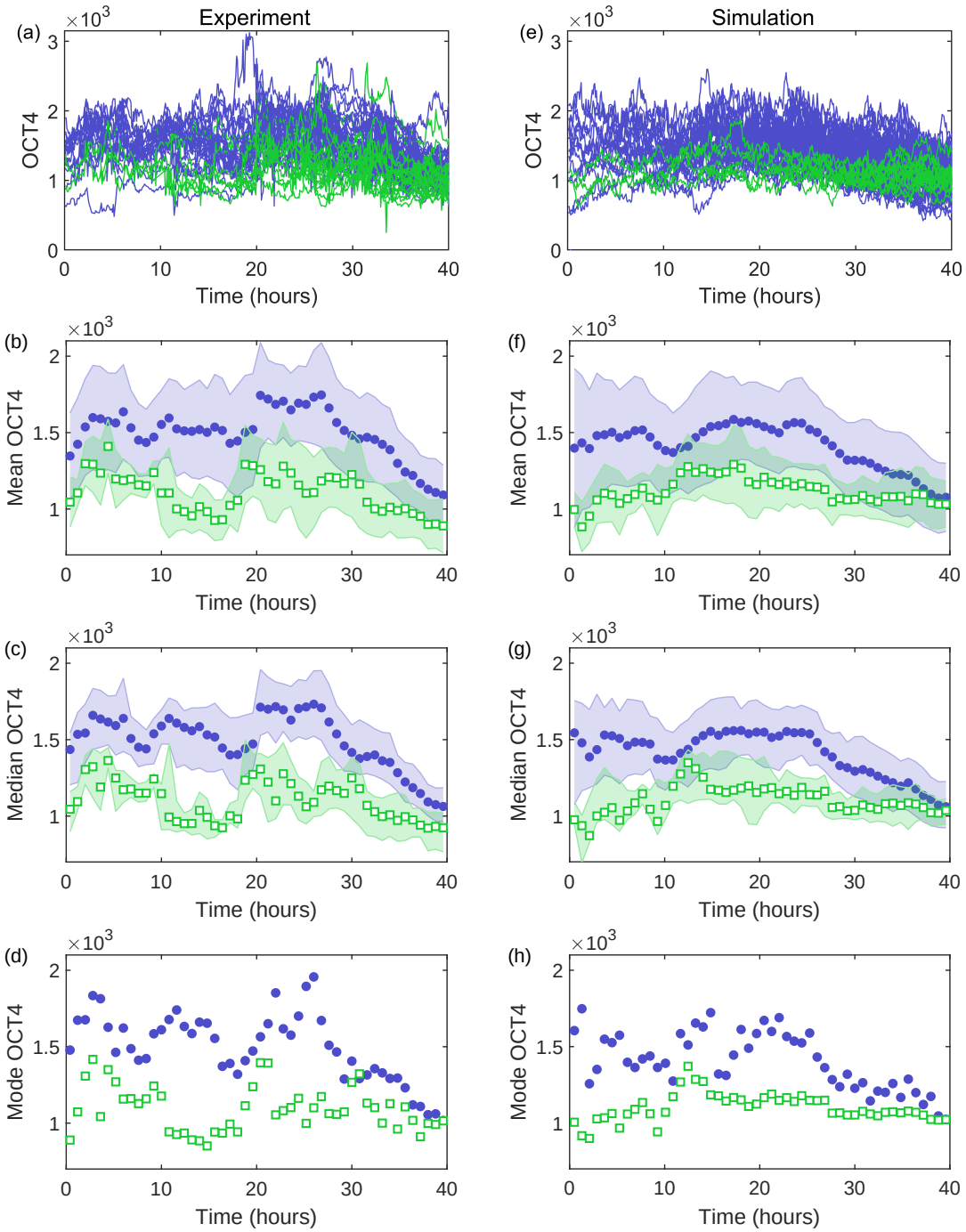

Figure 6.12: Dynamics for the (a-d) experimental OCT4 expression from Experiment 3 and (e-h) simulated OCT4 using the SLE with additive and multiplicative noise, and a time dependent carrying capacity, Eq (6.7), with parameters specified in Table 6.1.

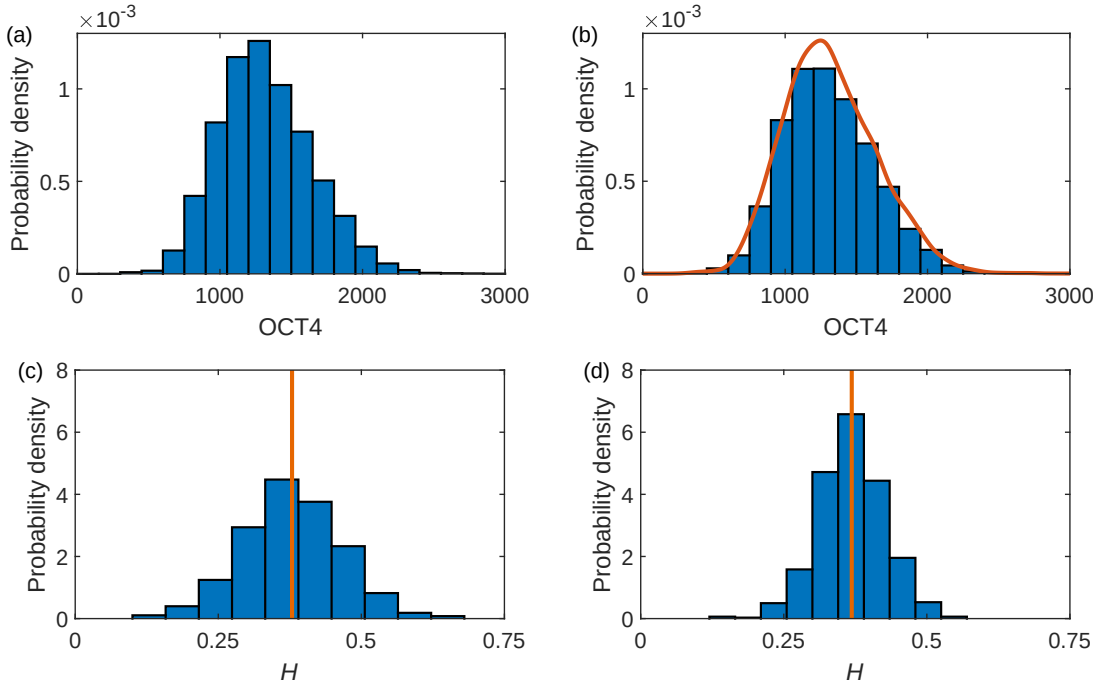

Figure 6.13: Distributions of the (a) experimental OCT4 expression from Experiment 3 and (b) simulated OCT4 (with the experimental distribution overlaid in orange) using the SLE with additive and multiplicative noise, and a time dependent carrying capacity, Eq (6.7), with parameters specified in Table 6.1. The Hurst exponent for all (c) experimental and (d) simulated cells. The orange lines show the mean Hurst exponent, $\overline{H} = 0.38$, for both experimental and simulated cells.

We can also consider the temporal distributions of OCT4 for each cell fate, shown for the experimental cells in Figure 6.9 and for the simulated cells in Figure 6.14. The model captures the shift to lower OCT4 values seen in pluripotent cells. The parameter choice could be further refined to additionally capture the change in the temporal skew (using time-dependent multiplicative noise as discussed in the previous section).

Here we have outlined some possible techniques for simulating temporal OCT4 using the SLE with different modes of fBm stochasticity and a time dependent carrying capacity. Note that we aim to illustrate the novel application of such a model and describe a framework which could be used to capture some of the global properties of experimental data sets. Further work is now required to elucidate the appropriate parameter choices and explore their biological implications.

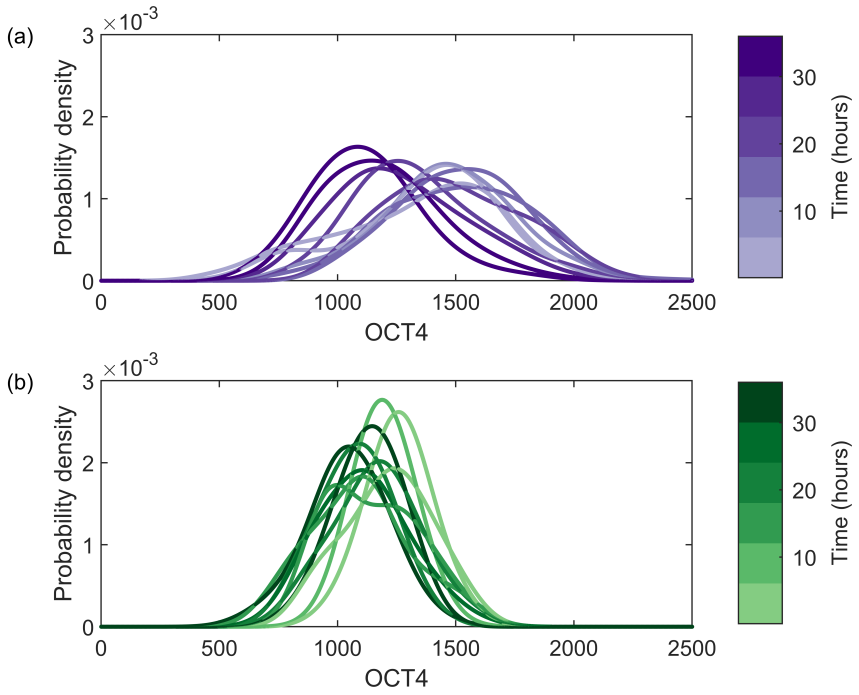

Figure 6.14: Temporal distributions of OCT4 for (a) pluripotent and (b) differentiated cells using a stochastic logistic equation with a shifting carrying capacity, Eq (6.7), with the parameters specified in Table 6.1.

6.2 Modelling cell differentiation

In the previous section we considered modelling temporal OCT4 regulation before any differentiation stimulus (BMP4) is added, corresponding to the time interval $0 < t < 40\,\text{h}$ in Experiment 3. The addition of BMP4 causes a significant reduction in OCT4 expression in the differentiated fate group of cells, shown in Figure 6.1. In this section we explore two methods of modelling this reduction in OCT4 as differentiation is induced. Firstly, we apply the SLE with a time-dependent carrying capacity as discussed previously, and secondly, we consider the use of the SLE with an Allee effect.

6.2.1 Differentiation with a time-dependent carrying capacity

We previously employed the SLE with a time-dependent carrying capacity, Eq (6.7), to simulate a moderate reduction in the average OCT4 expression post-25 hours, as shown in Figure 6.12. We could extend this technique to simulate the more drastic reduction in OCT4 seen when the differentiation stimulus is added, shown in Figure 6.1.

As before, we can estimate the carrying capacity for $t < 25\,\text{h}$ as $K_\text{p} \approx 1500$ and $K_\text{d} \approx 1100$ for pluripotent and differentiated cells, respectively. For $t > 25\,\text{h}$ we can simulate the modest reduction in OCT4 expression for the pluripotent cells with a reduction of

the carrying capacity, $K_\mathrm{p} \approx 1000$ (for simplicity we do not replicate this effect in the differentiated cells). For the differentiated cells in the time interval $t > 40\,\mathrm{h}$, $K_\mathrm{d} \approx 300$. These shifting carrying capacities, along with the other model parameters are given in Table 6.2. The dynamics under this regime are shown in Figure 6.15. The time-dependent carrying capacity leads to the reduction of OCT4 in the differentiated cell group, well capturing the dynamics of the experiment.

This model could be further refined by the use of a more sophisticated function for the time-dependent carrying capacity, which could be elucidated from the experimental data. However, the model is purely descriptive, with pluripotent and differentiated cells defined from the outset with different behavioural rules. Next we consider using the SLE with an Allee effect to simulate differentiation and identify the different cell fate types.

	Parameter	$0 \leq t < 25\,\mathrm{h}$	$25 \leq t < 40\,\mathrm{h}$	$40 \leq t < 68\,\mathrm{h}$
Pluripotent	N_0		14	
	r		0.015	
	K	1500	1000	1000
	σ_A		35	
	σ_M		0.035	
	H		0.38	
Differentiated	N_0		2	
	r	0.015	0.015	0.008
	K	1100	1100	300
	σ_A		25	
	σ_M		0.03	
	H		0.38	

Table 6.2: Simulation parameters for generating OCT4 expression for pluripotent and differentiated cells using the SLE with additive and multiplicative noise, and a time dependent carrying capacity, Eq (6.7), to capture induced differentiation.

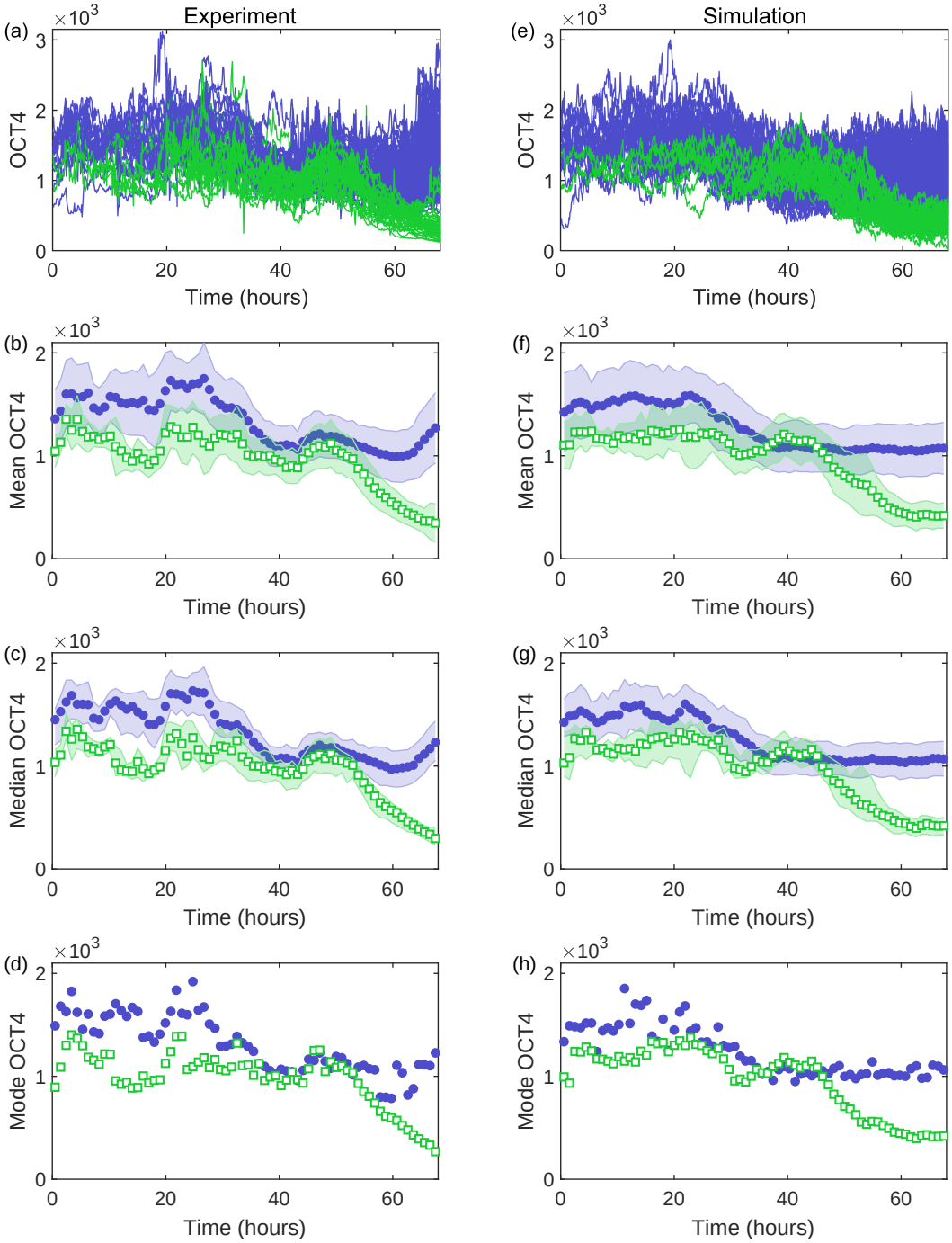

Figure 6.15: Dynamics for the (a-d) experimental OCT4 expression from Experiment 3 with differentiation induced at 40 hours and (e-h) simulated OCT4 using the SLE with additive and multiplicative noise, and a time dependent carrying capacity, Eq (6.7), with parameters specified in Table 6.2.

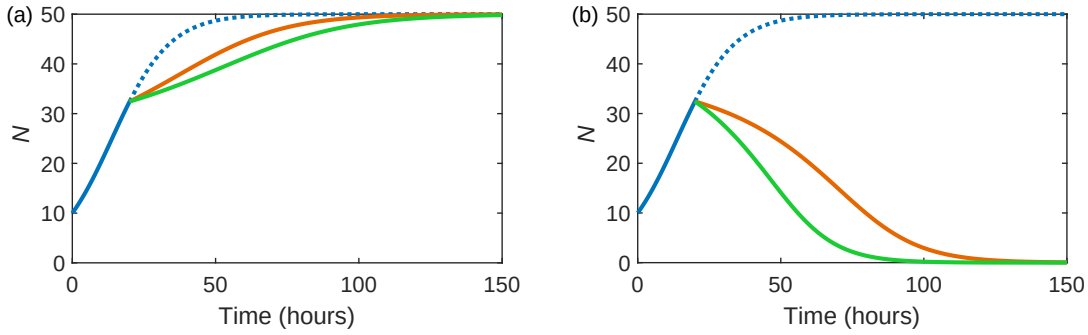

Figure 6.16: The deterministic logistic equation with an initial population size of $N_0 = 10$, $r = 0.1\,/\text{h}$ and $K = 50$ (blue). The Allee effect term in Eq (1.4) is introduced at $t = 25\,\text{h}$ with (a) $A = 20$ (orange) and $A = 25$ (green) and (b) $A = 40$ (orange) and $A = 50$ (green). The deterministic logistic growth with no Allee effect is shown as blue dashed.

6.2.2 Differentiation with an Allee effect

Another possible method of modelling induced differentiation is the SLE with a demographic Allee effect, Eq (1.4). Allee effects reduce population growth at low densities and are observed in both animal and cell populations [130–132]. Note that there are a range of methods for simulating Allee effects through e.g., difference equations [289, 290] and Lotka-Voltera models [252, 291]. Here we use the logistic equation and Eq (1.4) for consistency with our previous modelling results.

For a weak Allee effect ($A < N_0$), the change in population with time, dN/dt, remains positive for $N < K$ but is significantly suppressed. For a stonger Allee effect ($A > N_0$), dN/dt is negative for $N < K$ and results in the population declining to zero (presented in Section 1.3.2). It is this declining effect we can employ to simulate the reduction in OCT4 expression for the differentiated cells. The Allee effect term can be introduced at a certain time point during the logistic growth resulting in either continued suppressed growth or a decline to zero. Examples of 'switching on' both weak and strong Allee effects during logistic growth are shown in Figure 6.16.

For simulating OCT4 expression through the differentiation process with the SLE, we can switch on the Allee effect term at the time the differentiation agent is added (40 h). If the OCT4 expression is below A, then the Allee effect will be strong and the OCT4 will decline to zero. The stochasticity in the system will mean that only some of the cells will meet this condition, with others having an OCT4 expression greater than A, and therefore continuing with (suppressed) logistic growth. The stochasticity will also result in this effect taking place at all times past 40 h, so the differentiation process will happen for different cells at different times. The SLE for $X = \log(O)$ with additive fBm noise ξ_1

and multiplicative fBm noise ξ_2 is

$$\frac{dX}{dt} = r\left(1 - \frac{e^X}{K}\right)\left(\frac{O - A}{K}\right) + \frac{\sigma_{\mathrm{A}}}{e^X}\xi_1 + \sigma_{\mathrm{M}}\xi_2, \tag{6.8}$$

where A is the Allee effect critical point.

The OCT4 dynamics for 16 cells simulated with the SLE, Eq (6.6), for $t < 40\,\mathrm{h}$ and the SLE with an Allee effect, Eq (6.8), for $t \geq 40\,\mathrm{h}$ with $r = 0.025$, $K = 1290$, $\sigma_{\mathrm{A}} = 35$, $\sigma_{\mathrm{M}} = 0.035$ and $A = 1000$ are shown in Figure 6.17. Here the differentiated cells are identified at the end of the simulation as those cells whose OCT4 has reduced as a result of the Allee effect. The model captures the reduction of OCT4 in the differentiated subset of cells while keeping a remaining pluripotent cell population. However, the OCT4 in the differentiated group pre-Allee effect is no lower than for the pluripotent cell group, unlike in the experimental results. We also do not replicate the decline in OCT4 post-25 hours in the pluripotent cells.

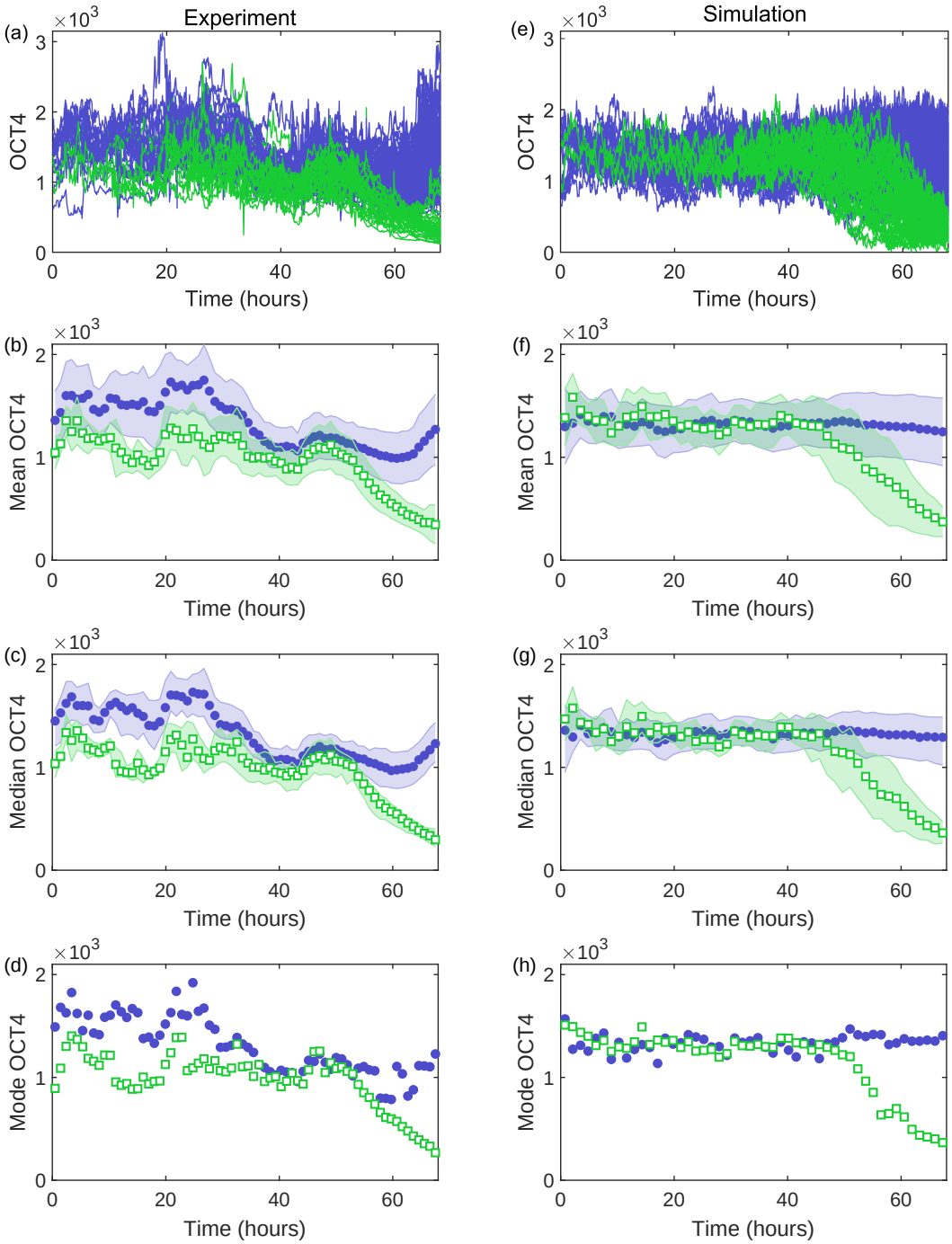

Figure 6.17: Dynamics for the (a-d) experimental OCT4 expression from Experiment 3 with differentiation induced at 40 hours and (e-h) simulated OCT4 using the SLE with an Allee effect at 40 hours, Eq (6.8), with $r = 0.025$, $K = 1290$, $\sigma_A = 35$, $\sigma_M = 0.035$ and $A = 1000$.

6.3 Discussion and conclusions

In this section we have explored different modelling techniques for describing the temporal OCT4 regulation observed in Experiment 3, particularly fractional Brownian motion and the stochastic logistic equation. A differentiation agent, BMP4, was added to the cells at 40 hours and results in the reduction of OCT4 expression in the differentiated cells, shown in Figure 6.1. All the models discussed follow a common base model which sets up the initial conditions and describes cell proliferation. When adjusted to produce unsynchronised cell divisions, the base model well describes the population growth with time, shown in Figure 6.3. We then focus on different mathematical methods of generating the temporal OCT4 expressions for the cell population within this base model.

Firstly, we consider modelling the OCT4 dynamics pre-BMP4, i.e., for $t < 40$ hours. The analysis in Chapter 5 revealed that OCT4 values fluctuate stochastically with anti-persistence and a Hurst exponent of 0.38, suggesting the use of fractional Brownian motion (fBm) [103]. There is also further experimental evidence that gene expressions and transcription factor dynamics display fractal characteristics [277]. The use of fBm is particularly common in financial modelling [292–294], but it has also been used to describe diffusion within crowded fluids (such as the cytoplasm of cells) [295], simulate stochastic textures and landscape topography [296, 297], and model coastal oil spills [298]. The stochasticity from fBm results in a wider range of OCT4 values at later times than seen experimentally (an effect which will become exacerbated with time).

The range of OCT4 can be controlled artificially with boundary conditions (either absorbing or reflecting), but the overall distribution of all OCT4 values is not well captured, shown in Figure 6.5. It is also unclear whether these boundary conditions are biologically appropriate as OCT4 expression is regulated by a complex range of factors across the transcriptional, post-transcriptional and epigenetic regulation levels [43, 242, 250, 263]. Interestingly, mechanical limits to transcription have been shown to naturally generate bounds to transcriptional noise [299]. A boundary condition at zero corresponds to the fact that OCT4 expression never becomes negative with the upper boundary representing a maximum possible value. Furthermore, when OCT4 expression reaches these boundaries, do cells die (i.e., are removed through absorbing boundaries) or does the OCT4 value recover (i.e., reflecting boundaries)? Although fBm alone is not sufficient to capture the experimental behaviour, it does (by design) capture the anti-persistence ($H = 0.38$) and so in all later model iterations we use fBm noise to generate the stochasticity.

A somewhat less artificial method of keeping the OCT4 values within range is to use the stochastic logistic equation (SLE), which has a regulating parameter of the carrying capacity, K. Classically this would represent the maximum amount of OCT4 that can be expressed within each individual cell, however the stochasticity allows for some

fluctuations above K. Similarly to the boundary conditions this maximum value will depend on the complex inter-regulatory network of OCT4, however, we estimate the value of the carrying capacity from the experimental results as the median of all OCT4 values (taking into account the stochasticity allowing for $O > K$). We consider both additive and multiplicative noise, shown in Figure 6.6 and Figure 6.8. The introduction of multiplicative noise creates larger fluctuations above the carrying capacity, qualitatively similar to those seen in the experiment. This results in a distribution of all OCT4 values well matched to the experiment, with the slight positive skew captured.

A property not captured by the SLE with either additive or multiplicative noise is the time-dependency of this positive skew. It occurs only at later times, and only in pluripotent cells, shown in the time-discretised distributions of OCT4 in Figure 6.9. This temporal skew can be captured by the SLE with both additive and multiplicative noise, with the type of noise time dependent; additive noise at early times produces symmetrical distributions of OCT4, with multiplicative noise at later times producing skewed distributions, shown in Figure 6.10. Here we changed the noise function stepwise, but this could be further smoothed using a more sophisticated time dependent noise function.

Another interesting property of the experimental OCT4 is the decline in expression for pluripotent cells post-25 hours, shown in Figure 6.11. We consider capturing this behaviour using the SLE with a time-dependent carrying capacity. Since this parameter is likely to depend on a myriad of biological factors, it is not unreasonable to expect that it may change with environmental conditions and experimental time. We consider the pluripotent and differentiated cells separately, each with a different carrying capacity, corresponding to the suggestion that the decision to differentiate is determined pre-differentiation stimulus [41]. The carrying capacity for both cell groups is reduced at 25 hours, resulting in a decline in OCT4 expression, particularly for the pluripotent cell group with originally higher expression. Although this technique well describes the experimental results (shown in Figure 6.12 and Figure 6.13), it requires multiple parameters which need to be elucidated from the experimental data.

We then consider modelling the OCT4 regulation for all times, including the decline in expression due to the addition of the differentiation stimulus. We extend the time-dependent carrying capacity method, reducing the carrying capacity further for the differentiated cell group at 40 hours. This well captures the decline in OCT4 upon differentiation, along with the more subtle decline in pluripotent cells, shown in Figure 6.15. Here we have used a stepwise change in the parameter K, but this is easily adjustable to other experimental results and more sophisticated functions could be used to capture other trends. The differentiated cells are identified from the outset and although this is not biologically unreasonable [41] the model itself does not produce the two fate groups, limiting its future capacity to develop into a predictive model.

A method of inducing differentiation which naturally produces the two fate groups is the SLE with an Allee effect. Allee effects are well used across mathematical biology [130–132], but this application to pluripotency transcription factor expression is novel. The Allee effect results in a decline to zero for cells whose OCT4 expression fluctuates below the critical point A. The stochasticity in the system means that this condition is met for only some of the cells, causing the formation of a differentiated cell group with reducing or zero OCT4 and a pluripotent cell group with stable OCT4 expression at the carrying capacity, shown in Figure 6.17. This model could be combined with a time-dependent carrying capacity to capture the decline in expression in pluripotent cells.

The models discussed here are of a purely descriptive nature, but outline a possible framework for modelling the regulation of OCT4. A summary of the models discussed and the experimental properties they capture is given in Table 6.3. Repeat experiments collecting the same data are needed to confirm which of these properties are inherent to OCT4 expression, and how they vary depending on experimental conditions. In general, this highlights the need for further temporal experimental data on PTF regulation to build upon this mathematical framework and develop more sophisticated predictive models.

	OCT4 behaviour				
	Anti-persistence	Distribution	Temporal skew	Shifting mean	Differentiation
Brownian motion					
fBm	✓				
SLE (additive fBm)	✓				
SLE (multiplicative fBm)	✓	✓			
SLE (add. & mult. fBm)	✓	✓	✓		
SLE (time-dependent K)	✓	✓	✓	✓	✓
SLE (Allee effect)	✓	✓	✓	✓	✓

Table 6.3: A summary of the models discussed to describe temporal OCT4 regulation and the behaviours captured by each.

Chapter 7

Future steps

In the preceding chapters we have quantified and modelled some key behaviours of hPSCs, including their kinematics, colony growth and aspects of intra-cellular pluripotency. In this chapter we discuss the next steps in the modelling of stem cell behaviours and iPSCs, stem cells which can be induced from other specialised cell types.

7.1 Deeper understanding of hPSCs

Throughout this thesis we have considered the kinematics, colony growth, and inter-cellular pluripotency of hPSCs and quantified, characterised, and explored methods of modelling these behaviours. Here we outline some of the opportunities for future research in these areas to further both the descriptive and predictive mathematical modelling of hPSCs.

7.1.1 Kinematics of hPSCs

In Chapters 2 and 3, we characterised the movements of individual and pairs of hPSCs. Using time-lapse microscopy imaging we quantified their division times, speeds, angles of movement, and diffusivities. There is clear scope to expand this analysis to cells in larger groups, continuing the bottom-up approach moving to triplets and quadruplets. We saw such larger groups of cells in Experiment 1, but these were not tracked. Example microscopy images of groups of cells from Experiment 1 are shown in Figure 7.1.

We found that, on average, individual cells within a pair move at a similar speed to individual cells ($18\,\mu$m/h) and that pairs as a whole have a similar diffusivity ($80\,\mu\text{m}^2$/h) to individual cells. Further experimental data of individual cell speeds belonging to different groups would be useful in clarifying if this is an inherent property of hPSCs. We would expect the migration speed and diffusivity to reduce when a cluster of cells reaches a certain size, as colonies qualitatively show reduced movement when compared to smaller groups. It would be interesting to measure both the collective migratory behaviour of colonies of different sizes and the speeds of individual cells within colonies to monitor the effect being part of a larger group of cells has on motility.

For the majority of the analysis of pairs of cells, we combined all pairs from Experiment 1 regardless of whether they originated from a cell division or were of 'unknown origin'. We are unable to sort the pairs of unknown origin into those from a cell division and those from individual cell migration as the pairs are seen together at the start of the image recording. It would be useful in further experiments to begin the imaging process earlier to facilitate

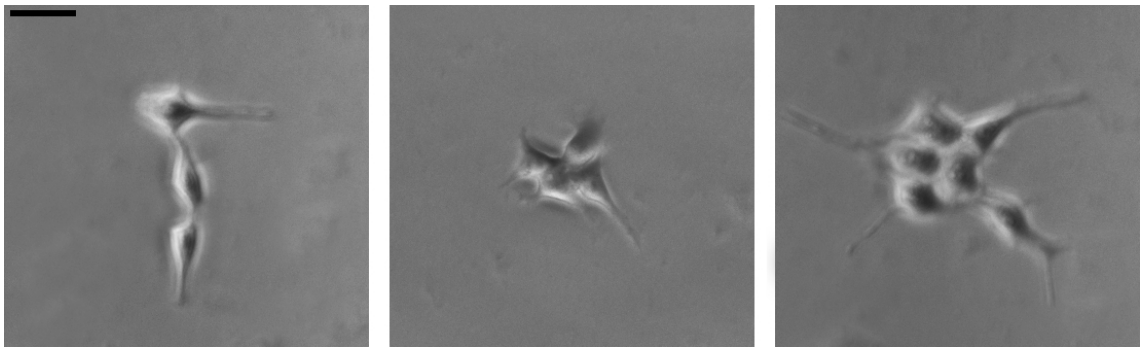

Figure 7.1: Microscopy images of groups of cells from Experiment 1. The scale bar represents 10 μm.

the identification and comparison of these two groups separately, and systematically. This would allow us to explore the effect clonality has on the cell pair behaviour. It would also be useful for analysis of colony clonality to identify what proportion of cell pairs we can expect to have been formed by individual cells which are not genetically identical. If pairs are formed in this way then the arising colonies will not be clonal. This data could also inform the clonality model discussed in Chapter 4.

Measuring the kinematic properties of cells from time-lapse imaging is challenging from an image analysis perspective. For Experiment 1, the cells were tracked manually, with their positions being recorded by hand for every image. Methods of automating the task still require a significant amount of manual calibration and work less effectively when considering groups of cells as it is often difficult to precisely identify the boundaries of individual cells. This itself is an avenue of future work, developing cell recognition software which can automate the manual tracking process from experimental images accurately. This would also allow for easier data collection and improved statistics in experiments such as these.

It is worth noting that here we consider only one experiment (Experiment 1) with 26 single cells and 50 pairs of cells. Although a large enough sample to perform statistical analysis, we do see a wide variation in cell behaviours and the collection of more data through further identical experimental repeats would be useful in validating the results.

7.1.2 Colony growth and clonality

In Chapter 4 we use experimental data of colony populations at 72 hours to infer the parameters for a stochastic exponential growth model. The model can be used to calculate the time at which clonality is lost due to merging colony populations. The model is a useful framework directly applicable to improving experimental protocols and also raises several important unanswered questions about the growth of colonies.

A barrier to the development of population models is the lack of experimental data

of population sizes over time for large colonies. The difficulties in data collection arise in automating the cell counting process during the image analysis stage. For our model, we use Experiment 2a which measures colony populations at a single time frame and use this to estimate the growth rate of the colonies. The collection of data through only one experiment is a major limitation of the results of this study and replicates need to be performed. It is also reasonable to expect that the growth rates of colonies might change over time and so experimental data at a larger number of time intervals would be useful in confirming our growth rate estimates.

There are also many other physical and biological factors which could affect the colony growth rates. We are only able to estimate the growth rates for single and pairs of cells upon seeding, further experimental data (of populations of groups of founder cells) are needed to elucidate the growth rates for larger clusters of cells. The growth rates could also be affected be the 'tightness' of the packing in each cell cluster. Since hPSCs proliferate more effectively in close proximity with each other, it is likely that a triplet of cells in contact grows faster than a triplet of cells separated by some distance.

The model also applies the assumption that the originally seeded cells are randomly scattered within the growth area. From this random scattering the expected clusters are predicted. It would be interesting to see if this is truly the case when carried out practically. This could then be compared to results using an automatic cell counter for seeding. Regardless of the results, experimental data on the clustering of cells after seeding would be easily implemented into the model as initial conditions. Secondarily, the packing of clusters could also be monitored to elucidate how the structures are most commonly formed.

There are also key links to be explored between the population model, the kinematic results in Chapters 2 and 3, and the future paths of exploration for the kinematics of cells described above. The migration speeds and diffusivities have knock-on effects on the clonal colony growth and once the super-diffusive properties of individual cells are understood more clearly, these individual movements can also be considered in the growing colonies simulation as a further effect through which clonality can be compromised.

It is worth noting that although there are many biological unknowns which are yet to be elucidated confidently from experimental results, the model provides a framework, easily adaptable to any of the possible experimental outcomes. It can be refined continuously and flexibly as more experimental data become available. Future developments of this model could be published as an interactive 'toolbox' for biologists, through which they could adjust the parameters to match their experimental protocols. There are similar tools under development to guide model construction for complex systems [97] and tissue simulators through which already established modelling frameworks can be explored [107].

Another method of simulating colony growth is agent-based modelling, where the

placement of each individual cell (or groups of cells) within the colony is considered. This commonly uses a cellular Potts approach [173, 175]. There is some evidence that cells divide less deeper in colony interiors due to increased pressure, but this has not been fully quantified. The clonality model of Chapter 4 should be compared to cellular Potts based models to formulate when the inclusion on individual cells is essential to capture the colony growth and resulting clonality loss.

7.1.3 Pluripotency

In Chapter 5 and 6, we use available experimental data (Experiment 3) to quantify and model the intra-cellular regulation of the pluripotency transcription factor OCT4. Although the modelling conditions are flexible, the analysis results of this section are limited by their basis on a single experiment. Furthermore, we consider OCT4 in isolation when in reality it is a part of a complex network of transcription factors which all interact to control cellular pluripotency.

Firstly, it would be informative to apply the same quantitative framework to the other predominant transcription factors, SOX2 and NANOG. Their individual regulatory dynamics could then be compared using the key descriptive parameters, and any systematic differences identified. This information will help build the picture of the wider PTF system. Next, the dynamics of the PTFs should be considered as part of an inter-linked network. Ideally, this would require the real-time monitoring of three gene expression levels within multiple colonies – an experimental challenge.

Quantitative characterisation of the regulation of the whole PTF network would guide more complex mathematical models of pluripotency. The results from Experiment 3 do provide suggestions for the types of models required; for example the quantification of the systematic regulation of OCT4 has implications for the underlying stochastic chemistry and can be used to inform chemical models of PTF regulation, often based on the Hill equation. The inclusion of the other key PTFs will also need to be carefully considered, as this will greatly increase the complexity of the models.

We have explored corresponding models for temporal OCT4 regulation, including fractional Brownian motion and the stochastic logistic equation, but these remain purely descriptive. The experimental evidence of fractal characteristics in gene expressions and transcription factor dynamics [277] highlights the need for the continued use of fractal techniques. We present two possible methods of simulating differentiation, time-dependent carrying capacities and Allee effects, which should be tested on other experimental data sets, and for the dynamics of other PTFs. There is also further work to be done in elucidating the model parameters, and clarifying their biological interpretation. In general, these further model developments, along with corresponding experimental studies of PTF regulation with time are needed to create more

fundamental models.

There are also interesting collective features of pluripotency on the colony scale. We can see the result of this in Experiment 3, with differentiated cells (with low OCT4 levels) clustered around the colony edge. To investigate the spatial clustering of cell pluripotency and cell generations, we carried out an experiment (Experiment 4) monitoring the levels of CellTrace™, OCT4 and NANOG in hESC colonies at 72 h (details given in Appendix A). An example stained colony showing the expression levels of CellTrace™ and NANOG is shown in Figure 7.2. A preliminary analysis of this experiment is given in Appendix C. The results suggest there is suppressed division in the colony centre, and tighter regulation in OCT4 and NANOG in larger colonies. Future work in this area will require more experiments, similar to Experiment 4, but with larger colonies to fully allow the spatial distributions to be explored. A deeper understanding of the collective effects of pluripotency and how cells divide within a colony will not only help to advance self-consistent mathematical models of stem cell colonies, but can also guide experimental protocols, particularly for isolating pluripotent parts of colonies and guiding controlled differentiation.

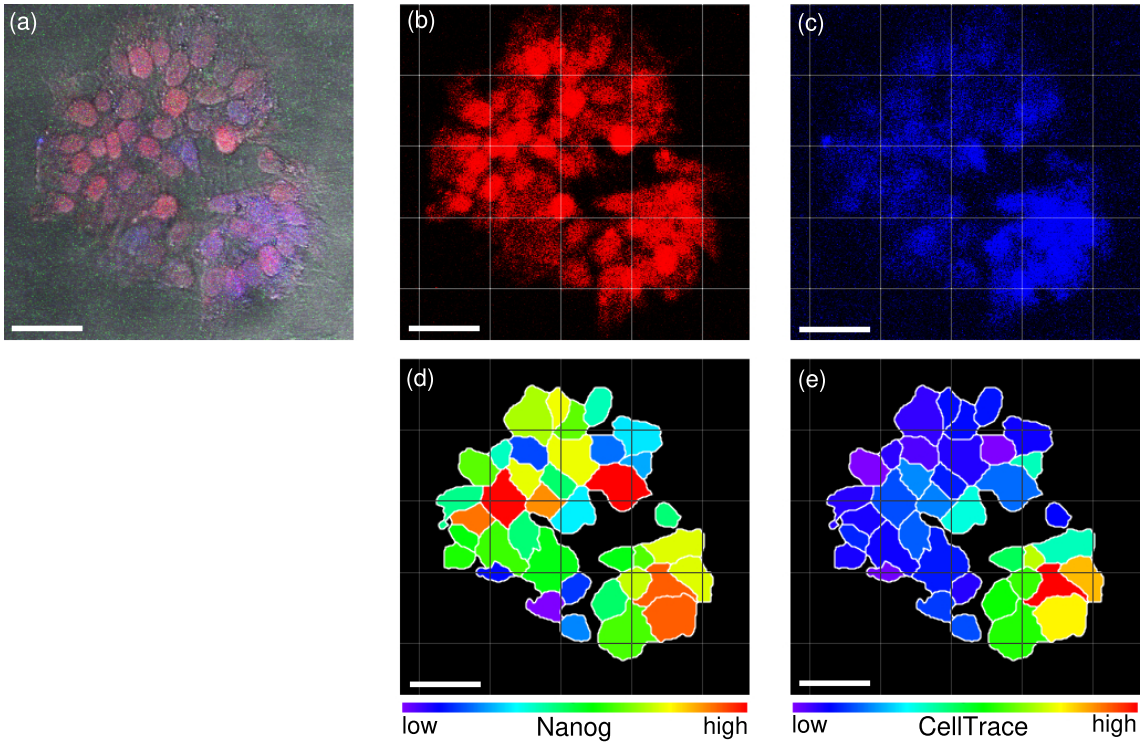

Figure 7.2: Example colony from Experiment 4. (a) The original colony microscopy images with the NANOG (red) and CellTrace™ (blue) staining visible. (b) NANOG and (c) CellTrace™ isolated. Image analysis allows the levels of expression of each staining to be quantified on a scale, shown for (d) NANOG and (e) CellTrace™. The scale bar represents 50 μm.

7.2 iPSCs

In many respects, iPSCs are the future of stem cell experiments; they are derived by the genetic reprogramming of differentiated cells [17] and can be made from other readily available specialised cell tissues. This has several advantages over traditional embryonic stem cells: firstly, iPSCs help to avoid the ethical issues associated with hESCs [19], and secondly, they offer patient-specific hPSCs which reduces the risk of transplanted tissue rejection [18]. However, the reprogramming of somatic cells into iPSCs is still a low-yield process in the laboratory [17].

A combination of experimental effort and computational modelling has shown that iPSC reprogramming is a continuous stochastic process and so most mathematical models in this area are probabilistic [83]. Some models directly consider the genetic network involved in the reprogramming, and some include the epigenetic network [300]. A recent model has described cell types as a set of hierarchically related dynamical attractors representing cell cycles [95]. Detailed reviews of the cell fate and reprogramming models can be found in [140] and [301].

The complexity of the system and the lack of appropriate experimental data that could be used to constrain such models appear to be the key challenges in the model development, verification and testing. Nevertheless, mathematical modelling has already revealed interesting behaviours in the reprogramming dynamics and a quantitative understanding of cell state transitions will facilitate the formulation of improved reprogramming strategies and guide the optimisation of reprogramming experiments.

Although the experiments and models discussed throughout this thesis have been constructed using hESC data, the mathematical framework is adaptable to iPSCs. The focus for iPSC modelling is the reprogramming process with many other general behaviours, such as cell migration and colony proliferation, yet to be fully explored. The existing experience with hPSC modelling is clearly relevant but more experimental work is needed to quantify the similarities and differences between the iPSCs, hESCs and other cell types.

Chapter 8

Summary and conclusions

Human pluripotent stem cells, hPSCs, are at the forefront of modern biological research due to their ability to differentiate into a wide range of cell types and self-renew through repeated divisions. However, they exhibit complex behaviour over multiple scales, and the *in-vitro* control of their differentiation trajectories is challenging. This thesis focuses on deepening our understanding of some of the key behaviours of hPSCs, through experimental analysis, rigorous quantification, and mathematical modelling.

Firstly, we use experimental time-lapse image analysis (Experiment 1) to extract key migratory properties of isolated and pairs of cells, quantifying their movements within the random walk framework. We then consider colony formation from a single cell and develop a stochastic exponential growth model (with the parameters estimated from Experiment 2) which can be used to explore the impact of colony growth on clonality loss. Finally, we consider the temporal regulation of intra-cellular pluripotency through the transcription factor OCT4 (Experiment 3), quantifying the regulatory properties and exploring several mathematical models through which different aspects of the behaviour can be described.

Below we summarise the main findings presented in the preceding chapters and the final conclusions which can be drawn from this work.

Behaviour of isolated hPSCs

Presented in Chapter 2, [88] and [89].

1. Movement can be described as an isotropic random walk with characteristic diffusivity at early times (up to 7 hours for unstained cells).
2. Quantitative parameters of individual cell movement extracted including: migration speeds, step lengths, time to division and angles of movement.
3. CellTrace™ reduces cell movement and cell survival.
4. Turning angles are biased towards zero and π creating a preferred axis of movement.
5. Cell motion is aligned with morphological elongation.
6. Morphological elongation of a cell is positively correlated with speed.

Behaviour of pairs of hPSCs

Presented in Chapter 3, [88] and [89].

1. Pair centroid movement can be described as an isotropic random walk with characteristic diffusivity at early times (up to 12 hours for unstained cells).
2. Quantitative parameters of individual cell movement extracted including: migration speeds, step lengths, cell separations, and angles of movement.
3. CellTrace™ reduces cell movement (both relative and centroid) and cell survival.
4. There is a preferred separation distance for pairs of approximately 70 μm.
5. The individual cells in a pair show the same preference for back-and-forth motion along the cell axis as isolated cells.
6. There is a preference for cells to move in the same direction as each other, possibly facilitated by the pseudopodia connection.

Growing colonies of hPSCs

Presented in Chapter 4 and [82].

1. Colony populations are multi-modal based on the initial number of founder cells.
2. Cell attachment rates to the substrate are quantified at 30–37%.
3. A stochastic growth model can be used to describe colony growth with stochastic growth rate parameters estimated from the experimental data.
4. From the model, the time at which single founder cell colonies are no longer identifiable based on colony size can be predicted.
5. Applying the model in space allows the prediction of the time clonality is lost at varying seeding densities.

Pluripotency

Presented in Chapter 5, 6 and [187].

1. OCT4 expression is correlated in sister cells and exacerbated upon BMP4 induction.
2. The correlation between sister OCT4 expression reduces over time.
3. The mean OCT4 expressions for each cell fate are quantified, with differentiated cells showing lower average OCT4.

4. The time-step differences in OCT4 can be described as a Laplace distribution.
5. OCT4 shows distinct anti-persistence (Hurst exponent 0.38).
6. The auto-correlation of OCT4 shows periods of anti-correlation (on average, 60% of a cell's lifetime).
7. Most auto-correlation functions can be described using $\cos(2\pi t/a)e^{-t/b}$ allowing the estimation of characteristic time-scale of the periodicity and the correlation decay time as $a = 12\,\mathrm{h}$ and $b = 3\,\mathrm{h}$, respectively.
8. The mean-square displacement of OCT4 expression is sub-diffusive, characterised by a power law up to 12 hours for pluripotent cells.
9. Fractional Brownian motion can be used to capture the anti-persistence, but alone is not enough to describe OCT4 regulation.
10. The stochastic logistic equation can be used to simulate OCT4 regulation, using both additive and multiplicative noise.
11. Differentiation could be simulated using the stochastic logistic equation with a time-dependent carrying capacity or an Allee effect.
12. Preliminary results suggest tighter regulation of pluripotency in large colonies, and spatial clustering of CellTrace™ and OCT4 in large colonies.

Throughout this thesis, we have applied well-known mathematical techniques to quantify and investigate some of the key biological properties of stem cells, particularly hPSCs. This mathematical framework can be transferred to other cell types, including iPSCs, and will be strengthened by its application to future experimental data. In Chapter 1, we highlighted the importance of an iterative interaction process between modelling and experimental work. It is worth noting that the work in each of these Chapters represents one iteration of this process. The next steps involve developing these models further through repeated comparison to future experiments.

The future goal is to grow this framework of techniques for modelling stem cells and produce open access 'toolbox' materials allowing biologists to adapt the models to their own experimental protocols. There are similar infrastructures under development in general for simulations of complex systems [97, 107]. A relevant note of the infrastructure in [97] is the requirement that models should be designed with a specific question in mind which allows them to be validated as 'fit for purpose'. With a wealth of available models of ever increasing complexity in stem cell biology, it is important to specify the problems we are trying to solve and take a focussed, systematic approach. In

a similar way to [97], we hope to take specific examples of stem cell behaviours and models to build techniques which can be more broadly applied and adapted.

The key findings above highlight the fruitfulness of an inter-disciplinary approach. It is crucial that the development of mathematical models is underpinned by the rigorous quantification of biological behaviours and experimental evidence. Through continued collaboration, we can develop non-invasive diagnostic tools, guide experimental protocol design and deepen our understanding of stem cell behaviours.

Appendix A

Experiment details

A.1 Experiment 1: time-lapse imaging of single and pairs of cells

Below are the details of the experimental methods for the Experiment 1 discussed in Chapters 2 and 3.

Human embryonic stem cell culture

Human embryonic stem cell line H9 at passages 40-42 (WiCell, Madison WI) was maintained on Matrigel® Basement Membrane Matrix (Corning Inc.) in mTeSR™1 medium (STEMCELL Technologies). Cells were dissociated into single-cell suspension with StemPro® Accutase® (Thermo Fisher) which was diluted by 50% with PBS and plated on 6-well plates pre-coated with Matrigel at a density of 1500 cells/cm^2 (as described in [87]) in mTeSR™1 media supplemented with 10 μM of Y-27632 (ROCK inhibitor, Chemdea) for the first five hours. Afterwards, cells were fed with TeSR™1 media and divided in two groups: in the absence and presence of Cell Tracer. The former group was incubated with 3 μM of DMSO for 20 minutes at 37°C. The latter group was stained with 3 μM CellTrace™ Violet Dye (Thermo Fisher) according to the manufacturer's protocol. Then, cells were washed twice with culture media before fresh mTeSR™1 was applied. After one hour cells were observed under time-lapse imaging.

Time-lapse video imaging and tracking

Cells were imaged using a Nikon Eclipse Ti-E microscope, with images recorded every 15 minutes over a total duration of 66 hours for the experiment in the absence of Cell Tracer. Each image, analysed within the Nikon NIS-Elements software, had a resolution of 0.62 and 0.96 μm/pixel, respectively. The sampling time of 15 min was chosen to ensure that the typical cell displacement between the images (46 μm) is several times the pixel size (such that the pixel discretisation has no significant effect on the trajectories and ensuring that the positions are not over-sampled) and several times smaller than the typical cell diameter (allowing each cell to be individually tracked and for the overall motion on scales larger than the cell size to be adequately sampled). Only cells which had no neighbours within the 150 μm radius were followed in view of the observations by Li *et al.* [87] that the mutual interaction of the cells is negligible at this and larger separations, thus preventing development of colonies arising from more than a single cell.

The single cells were tracked by manually recording their centroid coordinates in every frame. This approach is used extensively for cell tracking in general [302], as well as for hESCs [87]. For the single cell considered in Chapter 2, Section 2.2.4, the cell boundary and geometrical centre was tracked using ImageJ [303, 304]. Comparison of this to the previous coordinates taken by eye showed no significant difference and so we are confident that our results are robust to which of these tracking methods are used. Upon cell division, the daughter cells were tracked as a cell pair. Tracking of a single cell ceased when the cell underwent apoptosis; cell pairs were tracked until one of the cells underwent apoptosis or division, or one or both cells could no longer be clearly identified. The number of cells that died was also recorded, along with their time of death. We did not follow cell triples even when they were formed by division of a cell in a pair. Formation of a pair from convergence of two unrelated cells is rare since the individual random walks lead, on average, to the divergence of cell trajectories provided sufficient space is available.

The instantaneous velocity of a cell was obtained from its displacement between two consecutive frames. Circular statistics calculations were performed as described in Ref. [305] using Matlab and its Circular Statistics Toolbox (Directional Statistics) [206].

A.2 Experiment 2: colony populations and attachment to the substrate

Below are the details of the experimental methods for Experiment 2a and 2b discussed in Chapter 4.

Two types of experiments were carried out. Experiment 2a to collect data on colony numbers at 72 hours and Experiment 2b to collect data on the rates of cell attachment and the time to colony merging.

Experiment 2a: colony populations at 72 h

Human embryonic stem cells (WiCell, Madison WI) were plated at a density of 1500 cells/cm^2 onto 6-well plates coated with Matrigel® Basement Membrane Matrix (Corning Inc.), in the mTeSR1™ media (STEMCELL Technologies). The cells were stained with CellTrace™ Violet Dye (Thermo Fisher). At 72 hours after cell attachment the cells were fixed and microscopy images (Nikon Eclipse Ti-E microscope) were taken of the colonies. Data was then collected using Imaris Image Analysis Software (BITPLANE Inc) software to distinguish cell boundaries and count the number of cells in each colony. This data was extracted for 48 colonies.

Experiment 2b: cell attachment rates

Same set up as Experiment 2a. The numbers of cells attached 24 hours after seeding were recorded for different initial densities: 1000, 1200, 2000, 3000, 4000 and 7000 cells/cm^2. Microscopy images were also taken of these wells each day for eight days for the initial seeding densities 1200, 3000, 4000 and 7000 cells/cm^2.

A.3 Experiment 3: colony OCT4 expression

Below are the details of the experimental methods for Experiment 3 discussed in Chapter 5. This experiment was carried out by Purvis Lab (University of North Carolina, School of Medicine), and is published in [41]. Full experimental details and the experimental data set are available in [41].

Endogenous OCT4 levels in H9 wild-type cells and H9 OCT4-mCherry clone 8-2 were determined by antibody staining using a mouse antiOCT4 antibody (MABD76, EMD Millipore). Immunostaining was performed using standard protocols. Briefly, cells were fixed for 15 min in 4% paraformaldehyde and permeabilised and blocked for 30 min in 5% goat serum with 0.3% Triton X100 in TBS. Incubation with primary antibody was performed overnight, and the incubation with the secondary antibody (Molecular Probes) was done at room temperature for 45 min. Nuclei were visualized using NucBlue Fixed Cell Stain ready Probes reagent (R37606, Molecular Probes).

Asynchronous H9 OCT4-mCherry cells were plated on 12-well glass bottom plates (Cellvis) in phenol-red free or clear DMEM/F-12 (Gibco) supplemented with mTeSR1 supplement (05850, STEMCELL Technologies) approximately 24 h before being imaged. Cells were imaged using a Nikon Ti Eclipse microscope operated by NIS Elements software V4.30.02 with an Andor ZYLA 4.2 sCMOS camera and a custom stage enclosure (Okolabs) to ensure constant temperature, humidity, and CO_2 levels. Fresh media with or without BMP4 were added every 24 h. Images were flatfieldcorrected using NIS Elements.

A.4 Experiment 4: CellTrace™, OCT4 and Nanog

Below are the details of the experimental methods for Experiment 4 discussed in Chapter 7.

Human embryonic stem cells (WiCell, Madison WI) were plated at a density of 1500 cells/cm^2 onto 6-well plates coated with Matrigel® Basement Membrane Matrix (Corning Inc.), in the mTeSR1™ media (STEMCELL Technologies). The cells were

stained with CellTrace™ Violet Dye (Thermo Fisher) and OCT4 and Nanog markers. At 72 h after cell attachment the cells were fixed the and microscopy images (Nikon Eclipse Ti-E microscope) were taken of the colonies.

Data was then collected using Imaris Image Analysis Software (BITPLANE Inc) software to distinguish cell boundaries and count the number of cells in each colony. The fluorescent intensity levels of the three stains (CellTrace™, OCT4 and Nanog) were extracted in arbitrary fluorescent units (a.f.u) for every cell in every colony. Of the 48 colonies, 15 were stained with CellTrace™ only, 14 with additional OCT4 staining, and 19 with additional Nanog staining.

Appendix B

Circular statistics

When considering the directional migration of cells, the quantities of interest often include statistical properties (e.g., averages and standard deviations) of angular measures. The averaging of circular quantities (circular statistics), such as angles, has some peculiar properties. For example, consider taking the mean of the angles $\theta_1 = 0°$ and $\theta_2 = 270°$, shown in Figure B.1. By intuitively visualising the average direction of the two angles, the answer is $\overline{\theta} = 315°$. However, the arithmetic mean is $\overline{\theta} = (\theta_1 + \theta_2)/2 = 135°$, not the intuitive value or representative directional average in this case as it differs by 180° from the correct value $\theta = 315°$. The correct is given by the circular mean, defined for angles $\alpha_1, ..., \alpha_n$ as

$$\overline{\alpha} = \text{atan2}\left(\sum_{\text{j=1}}^{\text{n}} \sin\alpha_{\text{j}}, \sum_{\text{j=1}}^{\text{n}} \cos\alpha_{\text{j}}\right), \tag{B.1}$$

where arctan2 is a special trigonometric function, similar to the usual arctangent, except the signs of both arguments are used to determine the quadrant of the result,

$$\text{arctan2(y, x)} = \left\{ \begin{array}{ll} \arctan(\frac{y}{x}), & \text{if } x > 0 \\ \frac{\pi}{2} - \arctan(\frac{x}{y}) & \text{if } y > 0 \\ -\frac{\pi}{2} - \arctan(\frac{x}{y}) & \text{if } y < 0 \\ \arctan(\frac{y}{x}) \pm \pi & \text{if } x < 0 \\ \text{undefined} & \text{if } x = 0 \text{ and } y = 0 \end{array} \right\}. \tag{B.2}$$

Similarly, the circular correlation between two angular quantities is defined as

$$C = \frac{\sum_i \sin(\alpha_i - \overline{\alpha}) \sin(\beta_i - \overline{\beta})}{\sqrt{\sum_i \sin^2(\alpha_i - \overline{\alpha}) \sin^2(\beta_i - \overline{\beta})}}, \tag{B.3}$$

where α_i and β_i denote two samples of angles and $\overline{\alpha}$ and $\overline{\beta}$ their angular means. Further details of the use of circular statistics for biological applications can be found in Ref. [305]. Matlab users may also find the circular statistics toolbox (directional statistics) useful [206]. There are similar resources available for R [306].

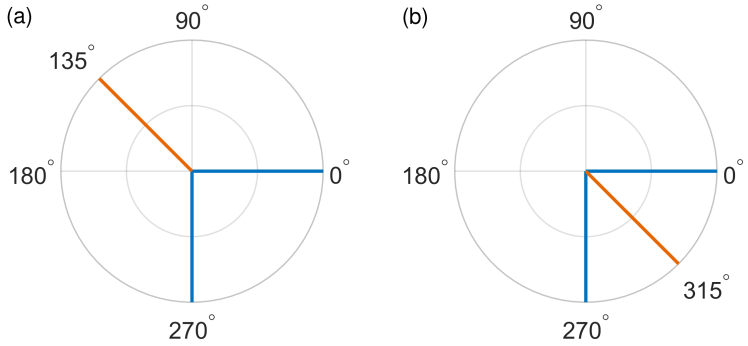

Figure B.1: Consider taking the average of two angles, $\theta_1 = 0°$ and $\theta_2 = 270°$ (blue). Using (a) the arithmetic mean gives $\overline{\theta} = 135°$ (orange) while using (b) the circular mean gives the more intuitive result $\overline{\theta} = 315°$ (orange).

Appendix C

Spatial structures: preliminary results

To further investigate the spatial clustering of cell pluripotency and cell generations, we carried out an experiment monitoring the levels of CellTrace™, OCT4 and NANOG in hESC colonies at 72 h.

Experiment 4 Carried out at the Institute of Biosciences at Newcastle University by Dr Irina Neganova. Data collection and image analysis by the author.

Single hESCs were plated at low density 1500 cells/cm^2. At 72 hours the cells were fixed (a preservation process) and their levels of CellTrace™ (generation marker), OCT4 and NANOG (pluripotency markers) were extracted using IMARIS image analysis software. All 48 colonies analysed were stained with CellTrace™. Of these, 15 were stained with CellTrace™ only, 14 with additional OCT4 staining, and 19 with additional NANOG staining. An example stained colony showing the expression levels of CellTrace™ and NANOG is shown in Figure 7.2.

CellTrace™ diagnostics are based on the dye dilution upon each cell division and therefore the extracted staining intensity at any later time is related to the initial amount of CellTrace™ at the start of the experiment, C_0. Since the amount of CellTrace™ should be conserved (just spread around a larger number of cells as more divisions occur), we can estimate C_0 as the sum of all the CellTrace™ values in the colony in the 72 h snapshot. Hence we scale CellTrace™ values by C_0 to make the colonies comparable. The conservation of the total marker amounts is not an issue for the pluripotency stains as they show the actively evolving level of gene expression in the cells (as in Experiment 3). We can summarise the levels of staining across the colonies using boxplots (Figure C.1) and histograms (Figure C.2).

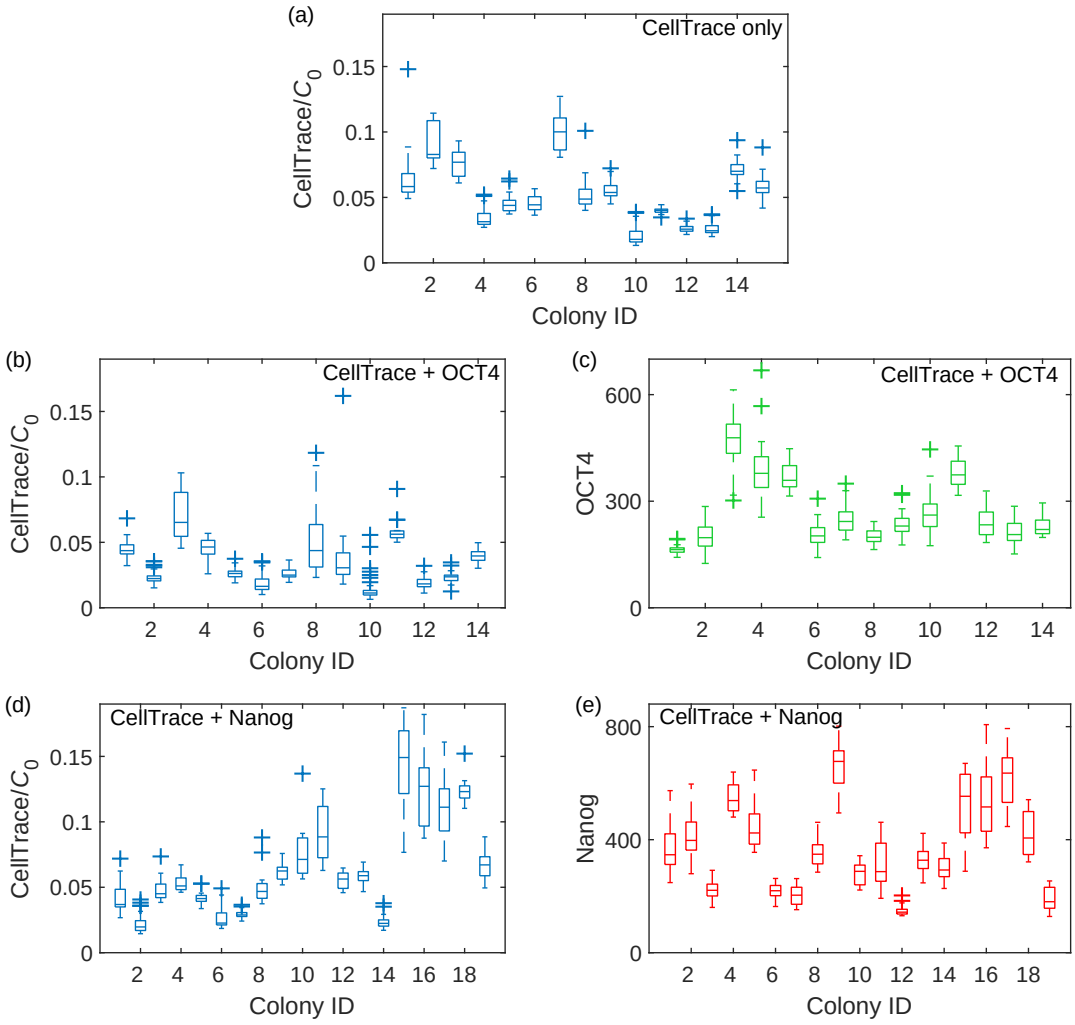

Figure C.1: Boxplots of the staining values for (a) CellTrace™ only, (b, c) CellTrace™ and OCT4 and (d, e) CellTrace™ and NANOG stained colonies from Experiment 4 at 72 h.

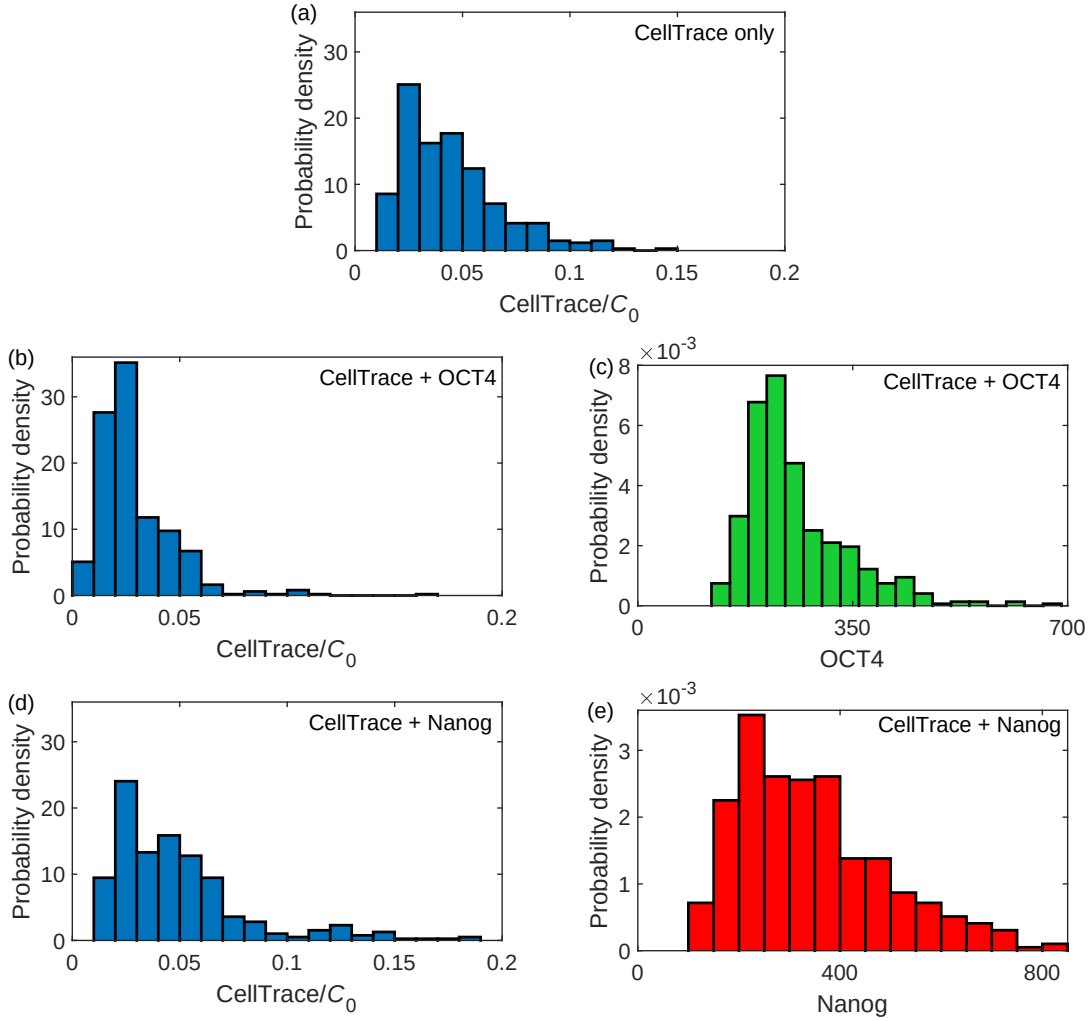

Figure C.2: Histograms of the staining values for (a) CellTrace™ only, (b, c) CellTrace™ and OCT4 and (d, e) CellTrace™ and NANOG stained colonies in Experiment 4 at 72 h.

The number of cell divisions a cell has undergone can be estimated from its current and initial CellTrace™ values, $C(72\,\text{h})$ and C_0, as $n = \log[C_0/C(72)]/\log(2)$ (as $C(72) = C_0/2^n$). The OCT4 and NANOG expressions for every cell in the experiment are shown for increasing n in Figure C.3. In Chapter 4 we identified a difference in growth rates between smaller (<30 cells) and larger (≥30 cells) colonies. In light of this the staining values are grouped by colony size in Figure C.3(a) and (b). This shows that on average, the cellular OCT4 value for cells in smaller colonies is higher and has greater variation than for larger colonies, with a mean (± standard deviation) OCT4 of 280 ± 113 and 245 ± 63 for smaller and larger colonies, respectively. This is the same for NANOG with a mean of 380 ± 162 and 293 ± 97 for smaller and larger colonies, respectively. This reduction in variation in larger colonies could be evidence of some collective regulation which occurs more strongly

either after a longer period of time, or in larger groups of cells.

Considering all cells together, there is a slight negative correlation between n and the pluripotency staining values, with a Pearson correlation coefficient ρ of -0.33 and -0.31 for OCT4 and NANOG, respectively. This suggests that for increasing number of divisions, there is some reduction in cellular pluripotency. We can also split the cells by colony, shown in Figure C.3(c) and (d). This shows that cells within the same colony that have undergone a similar number of divisions show similar pluripotency staining values.

We can examine the spatial variation of each staining, since it has been shown experimentally that differentiation occurs more readily at colony edges [61, 63, 187]. For every cell we consider its staining and its distance from the colony centroid position, r, shown in Figure C.4. Qualitatively, there is no visible correlation between the pluripotency stains and r (although there are very slight negative correlations suggested by the correlation coefficients with $\rho = -0.1$ and -0.2 for OCT4 and NANOG, respectively). For CellTrace™ there is a visible decline as the distance from the colony centre increases, with $\rho = -0.34$. This suggests that the cells which have undergone the most amount of divisions (and therefore have the least CellTrace™) are the furthest away from the colony centre. This could be evidence of suppressed division in the colony centre.

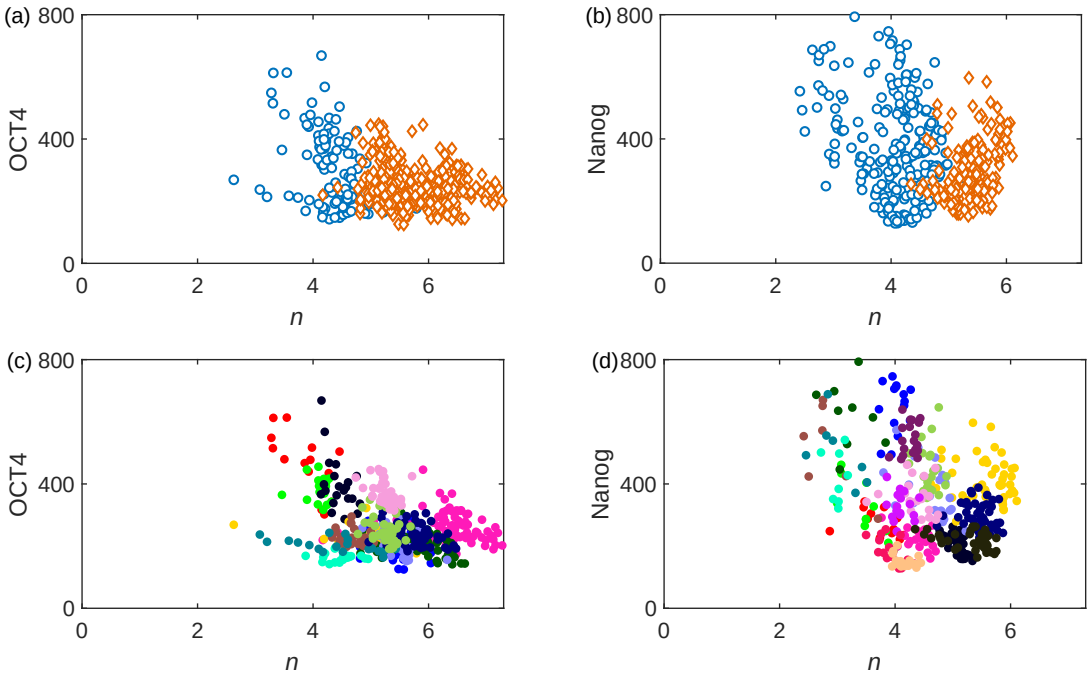

Figure C.3: The number of cell divisions and staining values for all cells for (a) OCT4 and (b) NANOG with smaller colonies (<30 cells) shown as blue circles and larger colonies (≥30 cells) as orange diamonds. Staining values for all cells with increasing n for (c) OCT4 and (d) NANOG. Each colour represents a separate colony.

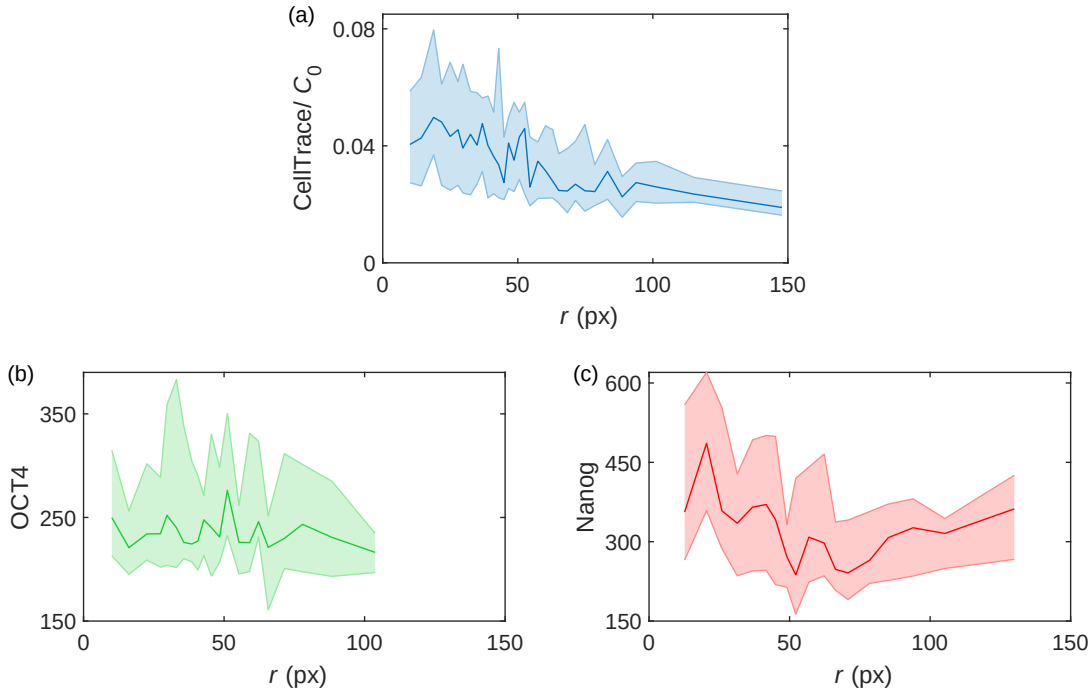

Figure C.4: The median staining values with inter-quartile ranges for increasing distance from the colony centroid, r, for (a) CellTrace™, (b) OCT4 and (c) NANOG.

The lack of obvious spatial correlation between the pluripotency staining could be due to the limitation of colony size. The largest colony in the experiment has 77 cells, with most of the colonies consisting of 20–30 cells. These colonies may not be large enough for the collective effects of pluripotency to either take effect, or become visible. We can consider the clustering of the staining in the largest colony, which is stained with CellTrace™ and OCT4, shown in Figure C.5. If we bin the cells according to their staining value as either below the colony mean ('low') or above the colony mean ('high') we can see a clear clustering in OCT4, with a division between the two groups through the middle of the colony, shown in Figure C.5(d). For the CellTrace™ staining, any clustering patterns are not apparent, as shown in Figure C.5(c).

We can explore this clustering further using spatial statistics techniques. One possibility is to quantify the clustering using Ripley's K function. For a point pattern Ripley's K function, $K(r)$, determines if particles are dispersed, clustered, or randomly distributed throughout the area contained by a search radius, r. It can be calculated by

$$K(r) = \lambda^{-1} \sum_{i \neq j} \frac{I(d_{ij} < r)}{n}, \tag{C.1}$$

where d_{ij} is the Euclidean distance between the i^{th} and j^{th} points in a set of n data points,

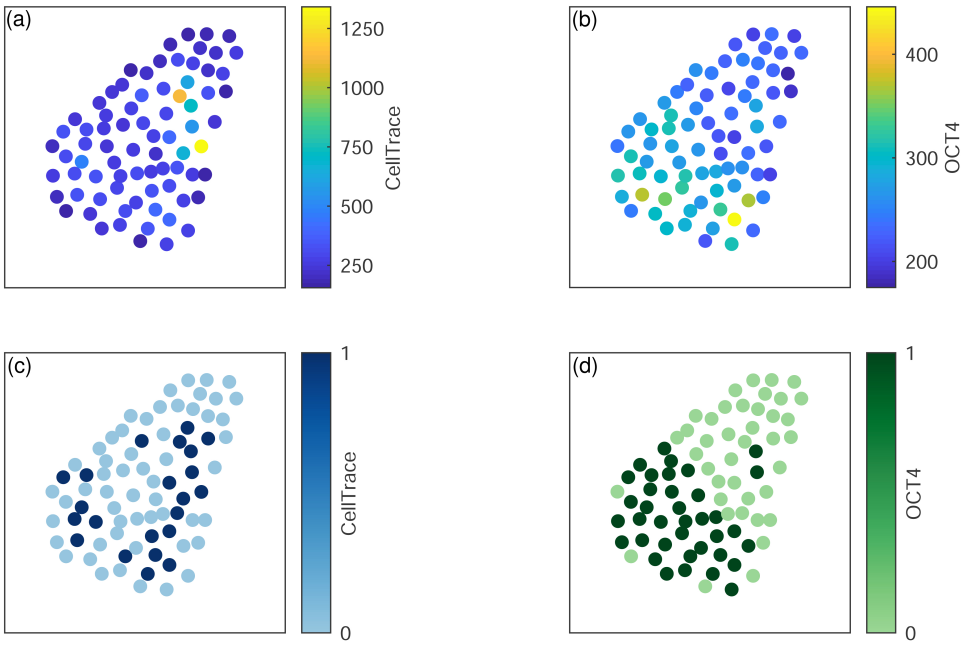

Figure C.5: The largest colony colour coded by (a) CellTrace™ and (b) OCT4 staining levels. In (c) and (d) cells are binned into 'high' (1) and 'low' (0) staining groups based on whether they fall above or below the colony mean for (c) CellTrace™ and (d) OCT4.

λ is the average density of points and I is the indicator function (1 if the operand is true, 0 if false). The function $K(r)$ is shown in Figure C.6 for CellTrace™ and OCT4, and for simulated samples of colonies with the same number of cells, but randomly allocated staining values. In both cases $K(r)$ is consistently above the expected values for randomly dispersed cells, suggesting that there is clear evidence of spatial clustering in both the CellTrace™ and OCT4 stains.

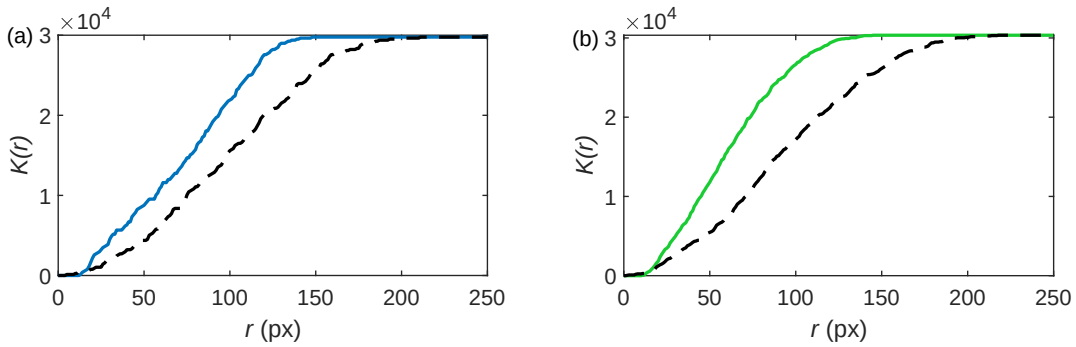

Figure C.6: Ripley's K function, $K(r)$, for (a) CellTrace™ and (b) OCT4. The black dashed lines show $K(r)$ for the same number of randomly scattered cells.

Bibliography

[1] D. G. Drubin and A. A. Hyman. Stem cells: the new "model organism". *Mol. Biol. Cell*, 28(11):1409–1411, 2017.

[2] C. E. Murry and G. Keller. Differentiation of embryonic stem cells to clinically relevant populations: Lessons from embryonic development. *Cell*, 132(4):661–680, 2008.

[3] J. M. W. Slack. *What is a stem cell?*, chapter 1, pages 1–11. John Wiley & Sons, Ltd, 2017.

[4] D. Orlic, J. Kajstura, S. Chimenti, I. Jakoniuk, S. M. Anderson, B. Li, J. Pickel, R. McKay, B. Nadal-Ginard, D. M. Bodine, A. Leri, and P. Anversa. Bone marrow cells regenerate infarcted myocardium. *Pediatr. Transplant.*, 7:86–88, 2015.

[5] K. Y. Helmy, S. A. Patel, K. Silverio, L. Pliner, and P. Rameshwar. Stem cells and regenerative medicine: accomplishments to date and future promise. *Ther. Deliv.*, 1:693–705, 2010.

[6] S. Bonner-Weir. Stem cells in diabetes: what has been achieved. *Horm. Res.*, 60:693–705, 2003.

[7] N. I. zur Nieden, G. Kempka, and H. J. Ahr. In vitro differentiation of embryonic stem cells into mineralized osteoblasts. *Differentiation*, 71:18–27, 2003.

[8] K. Schwanke, S. Wunderlich, M. Reppel, M. E. Winkle, M. Matzkies, S. Groos, J. Itskovitz-Eldor, A. R. Simon, J. Hescheler, A. Haverich, and U. Martin. Generation and characterization of functional cardiomyocytes from rhesus monkey embryonic stem cells. *Stem Cells*, 24:1423–1432, 2006.

[9] V. Tabar and L. Studer. Pluripotent stem cells in regenerative medicine: challenges and recent progress. *Nat. Rev. Genet.*, 15(2):82–92, 2014.

[10] H. T. Nguyen, K. Jacobs, and C. Spits. Human pluripotent stem cells in regenerative medicine: where do we stand? *Reproduction*, 156(5):143–153, 2018.

[11] A. Trounson and C. McDonald. Stem cell therapies in clinical trials: progress and challenges. *Cell Stem Cell*, 17(1):11–22, 2015.

[12] G. Shroff, J. Dhanda Titus, and R. Shroff. A review of the emerging potential therapy for neurological disorders: human embryonic stem cell therapy. *Am. J. Stem Cells*, 6(1):1–12, 2017.

[13] A. D. Ebert and C. N. Svendsen. Human stem cells and drug screening: opportunities and challenges. *Nat. Rev. Drug Discov.*, 9(5):367–372, 2010.

[14] Y. J. Lou and X. G. Liang. Embryonic stem cell application in drug discovery. *Acta Pharmacol. Sin.*, 32(2):152–159, 2011.

[15] E. K. Zuba-Surma, W. Wojakowski, Z. Madeja, and M. Z. Ratajczak. Stem cells as a novel tool for drug screening and treatment of degenerative diseases. *Curr. Pharm. Des.*, 18:2644–2656, 2012.

[16] Y. Avior, I. Sagi, and N. Benvenisty. Pluripotent stem cells in disease modelling and drug discovery. *Nat. Rev. Mol. Cell Biol.*, 17(3):170182, 2016.

[17] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5):861–872, 2007.

[18] B. Blum and N. Benvenisty. The tumorigenicity of human embryonic stem cells. *Adv. Cancer Res.*, 100:133–158, 2008.

[19] Zachary J. Kastenberg and Jon S. Odorico. Alternative sources of pluripotency: science, ethics, and stem cells. *Transplant. Rev.*, 22(3):215–222, 2008.

[20] C. Kropp, D. Massai, and R. Zweigerdt. Progress and challenges in large-scale expansion of human pluripotent stem cells. *Process Biochem.*, 59:244–254, 2017.

[21] I. Barbaric, V. Biga, P. J. Gokhale, M. Jones, D. Stavish, A. Glen, D. Coca, and P. W. Andrews. Time-lapse analysis of human embryonic stem cells reveals multiple bottlenecks restricting colony formation and their relief upon culture adaptation. *Stem Cell Rep.*, 3:142–155, 2014.

[22] K. Watanabe, M. Ueno, D. Kamiya, A. Nishiyama, M. Matsumura, T. Wataya, J. B. Takahashi, S. Nishikawa, S. Nishikawa, K. Muguruma, and Y. Sasai. A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.*, 25(6):681–686, 2007.

[23] K. Tokunaga, N. Saitoh, I. G. Goldberg, C. Sakamoto, Y. Yasuda, Y. Yoshida, S. Yamanaka, and M. Nakao. Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells. *Sci. Rep.*, 4:6996, 2014.

[24] M. Maddah, U. Shoukat-Mumtaz, S. Nassirpour, and K. Loewke. A system for automated, noninvasive, morphology-based evaluation of induced pluripotent stem cell cultures. *J. Lab. Autom.*, 19(5):454–460, 2014.

[25] M. Suga, H. Kii, K. Niikura, Y. Kiyota, and M. K. Furue. Development of a monitoring method for non-labeled human pluripotent stem cell growth by time-lapse image analysis. *Stem Cells Transl. Med.*, 4(7):720–730, 2015.

[26] K. A. Becker, P. N. Ghule, J. A. Therrien, J. B. Lian, J. L. Stein, A. J. van Wijnen, and G. S. Stein. Self-renewal of human embryonic stem cells is supported by a shortened G1 cell cycle phase. *J. Cell. Physiol.*, 209(3):883–893, 2006.

[27] S. Pauklin and L. Vallier. The cell-cycle state of stem cells determines cell fate propensity. *Cell*, 156(6):1338, 2014.

[28] S. Dalton. Linking the Cell Cycle to Cell Fate Decisions. *Trends Cell Biol.*, 25(10):592–600, 2015.

[29] P. W. Andrews N. J. Harrison, D. Baker. Culture adaptation of embryonic stem cells echoes germ cell malignancy. *Int. J. Androll*, 30:275–281, 2007.

[30] Y. B. Park, Y. Y. Kim, S. K. Oh, S. G. Chung, S. Y. Ku, S. H. Kim, Y. M. Choi, and S. Y. Moon. Alterations of proliferative and differentiation potentials of human embryonic stem cells during long-term culture. *Exp. Mol. Med.*, 40(1):98–108, 2008.

[31] D. A. Claassen, M. M. Desler, and A. Rizzino. ROCK inhibition enhances the recovery and growth of cryopreserved human embryonic stem cells and human induced pluripotent stem cells. *Mol. Reprod. Dev.*, 76(8):722–732, 2009.

[32] J. Turner, L. Quek, D. Titmarsh, J. O. Krmer, L. Kao, L. Nielsen, E. Wolvetang, and J. Cooper-White. Metabolic profiling and flux analysis of MEL-2 human embryonic stem cells during exponential growth at physiological and atmospheric oxygen concentrations. *PLoS One*, 9(11):1–13, 2014.

[33] P. N. Ghule, R. Medina, C. J. Lengner, M. Mandeville, M. Qiao, Z. Dominski, J. B. Lian, J. L. Stein, A. J. van Wijnen, and G. S. Stein. Reprogramming the pluripotent cell cycle: Restoration of an abbreviated G1 phase in human induced pluripotent stem (iPS) cells. *J. Cell. Physiol.*, 226(5):1149–1156, 2011.

[34] G. Dravid, Z. Ye, H. Hammond, G. Chen, A. Pyle, P. Donovan, X. Yu, and L. Cheng. Defining the role of Wnt/β-Catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells. *Stem Cells*, 23(10):1489–1501, 2005.

[35] H. Qin, T. Yu, T. Qing, Y. Liu, Y. Zhao, J. Cai, J. Li, Z. Song, X. Qu, P. Zhou, J. Wu, M. Ding, and H. Deng. Regulation of apoptosis and differentiation by p53 in human embryonic stem cells. *J. Biol. Chem.*, 282(8):5842–5852, 2007.

[36] T. TeSlaa, K. Setoguchi, and M. A. Teitell. Mitochondria in human pluripotent stem cell apoptosis. In *Seminars in cell & developmental biology*, volume 52, pages 76–83. Elsevier, 2016.

[37] K. Setoguchi, T. TeSlaa, C. M. Koehler, and M. A. Teitell. P53 regulates rapid apoptosis in human pluripotent stem cells. *J. Mol. Biol.*, 428(7):1465–1475, 2016.

[38] L. Romorini, X. Garate, G. Neiman, C. Luzzani, V. A. Furmento, A. S. Guberman, G. E. Sevlever, M. E. Scassa, and S. G. Miriuka. AKT/GSK3β signaling pathway is critically involved in human pluripotent stem cell survival. *Sci. Rep.*, 6:35660, 2016.

[39] M. Li and J. C. Izpisua Belmonte. Deconstructing the pluripotency gene regulatory network. *Nat. Cell Biol.*, 20(4):382–392, 2018.

[40] X. Zhang, I. Neganova, S. Przyborski, C. Yang, M. Cooke, S. P. Atkinson, G. Anyfantis, S. Fenyk, W. N. Keith, S. F. Hoare, O. Hughes, T. Strachan, M. Stojkovic, P. W. Hinds, L. Armstrong, and M. Lako. A role for NANOG in G1 to S transition in human embryonic stem cells through direct binding of CDK6 and CDC25A. *J. Cell Biol.*, 184(1):67–82, 2009.

[41] S. C. Wolff, K. M. Kedziora, R. Dumitru, C. D. Dungee, T. M. Zikry, A. S. Beltran, R. A. Haggerty, J. Cheng, M. A. Redick, and J. E. Purvis. Inheritance of OCT4 predetermines fate choice in human embryonic stem cells. *Mol. Syst. Biol.*, 14(9):e8140, 2018.

[42] R. M. Kumar, P. Cahan, A. K. Shalek, R. Satija, A. DaleyKeyser, H. Li, J. Zhang, K. Pardee, D. Gennert, J. J. Trombetta, T. C. Ferrante, A. Regev, G. Q. Daley, and J. J. Collins. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature*, 516(7529):56–61, 2014.

[43] Z. Wang, E. Oron, B. Nelson, S. Razis, and N. Ivanova. Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells. *Cell Stem Cell*, 10(4):440–454, 2012.

[44] J. Lee, Y. Go, I. Kang, Y. Han, and J. Kim. Oct-4 controls cell-cycle progression of embryonic stem cells. *Biochem. J.*, 426(2):171–181, 2010.

[45] J. Ouyang, W. Yu, J. Liu, N. Zhang, L. Florens, J. Chen, H. Liu, M. Washburn, D. Pei, and T. Xie. Cyclin-dependent kinase-mediated Sox2 phosphorylation enhances the ability of Sox2 to establish the pluripotent state. *J. Biol. Chem.*, 290(37):22782–22794, 2015.

[46] G. S. Belinsky and S. D. Antic. Mild hypothermia inhibits differentiation of human embryonic and induced pluripotent stem cells. *BioTechniques*, 55(2):79–82, 2013.

[47] L. Kent. Culture and maintenance of human embryonic stem cells. *J. Vis. Exp.*, page e1427, 2009.

[48] N. S. Hwang, S. Varghese, and J. Elisseeff. Controlled differentiation of stem cells. *Adv. Drug Deliv. Rev.*, 60(2):199–214, 2008.

[49] M. H. Zaman, L. M. Trapani, A. L. Sieminski, D. MacKellar, H. Gong, R. D. Kamm, A. Wells, D. A. Lauffenburger, and P. Matsudaira. Migration of tumor cells in 3D matrices is governed by matrix stiffness along with cell-matrix adhesion and proteolysis. *Proc. Natl. Acad. Sci. U.S.A.*, 103(29):10889–10894, 2006.

[50] T. Vazin and W. J. Freed. Human embryonic stem cells: derivation, culture, and differentiation: a review. *Restor. Neurol. Neurosci.*, 28(4):589–603, 2010.

[51] C. L. Bauwens, R. Peerani, S. Niebruegge, K. A. Woodhouse, E. Kumacheva, M. Husain, and P. W. Zandstra. Control of human embryonic stem cell colony and aggregate size heterogeneity influences differentiation trajectories. *Stem Cells*, 26:2300–2310, 2008.

[52] G. Chen, Z. Hou, D. R. Gulbranson, and J. A. Thomson. Actin-myosin contractility is responsible for the reduced viability of dissociated human embryonic stem cells. *Cell Stem Cell*, 7(2):240–248, 2010.

[53] R. J. Petrie, A. D. Doyle, and K. M. Yamada. The physics of adherent cells. *Rev. Mod. Phys.*, 85(3):1327–1381, 2013.

[54] N. Heins, A. Lindahl, U. Karlsson, M. Rehnstrm, G. Caisander, K. Emanuelsson, C. Hanson, H. Semb, P. Bjrquist, P. Sartipy, and J. Hyllner. Clonal derivation and characterization of human embryonic stem cell lines. *J. Biotechnol.*, 122(4):511–520, 2006.

[55] I. Glauche, L. Bystrykh, C. Eaves, and I. Roeder. Stem cell clonality - theoretical concepts, experimental techniques, and clinical challenges. *Blood Cell. Mol. Dis.*, 50(4):232–240, 2013.

[56] J. Chang, M. H. Kim, E. Agung, S. Senda, and M. Kino-Oka. Effect of migratory behaviors on human induced pluripotent stem cell colony formation on different extracellular matrix proteins. *Regen. Ther.*, 10:27–35, 2019.

[57] E. Shuzui, M. Kim, and M. Kino-oka. Anomalous cell migration triggers a switch to deviation from the undifferentiated state in colonies of human induced pluripotent stems on feeder layers. *J. Biosci. Bioeng.*, 127(2):246–255, 2019.

[58] J. Wu, Y. Fan, and E. S. Tzanakakis. Increased culture density is linked to decelerated proliferation, prolonged G1 phase, and enhanced propensity for differentiation of self-renewing human pluripotent stem cells. *Stem Cells Dev.*, 24(7):892–903, 2015.

[59] A. Nemashkalo, A. Ruzo, I. Heemskerk, and A. Warmflash. Morphogen and community effects determine cell fates in response to BMP4 signaling in human embryonic stem cells. *Development*, 144(17):3042–3053, 2017.

[60] B. D. MacArthur and I. R. Lemischka. Statistical mechanics of pluripotency. *Cell*, 154:484–489, 2013.

[61] K. A. Rosowski, A. F. Mertz, S. Norcross, E. R. Dufresne, and V. Horsley. Edges of human embryonic stem cell colonies display distinct mechanical properties and differentiation potential. *Sci. Rep.*, 5:14218, 2015.

[62] B. R. Gorman, J. Lu, A. Baccei, N. C. Lowry, J. E. Purvis, R. S. Mangoubi, and P. H. Lerou. Multi-scale imaging and informatics pipeline for in situ pluripotent stem cell analysis. *PLoS One*, 9, 2014.

[63] A. Warmflash, B. Sorre, F. Etoc, E. D. Siggia, and A. H. Brivanlou. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods*, 11(8):847–854, 2014.

[64] F. Etoc, J. Metzger, A. Ruzo, C. Kirst, A. Yoney, M. Z. Ozair, A. H. Brivanlou, and E. D. Siggia. A balance between secreted inhibitors and edge sensing controls gastruloid self-organization. *Dev. Cell*, 39(3):302–315, 2016.

[65] X. Xue, Y. Sun, A. M. Resto-Irizarry, Y. Yuan, K. M. Aw Yong, Y. Zheng, S. Weng, Y. Shao, Y. Chai, L. Studer, and J. Fu. Mechanics-guided embryonic patterning of neuroectoderm tissue from human pluripotent stem cells. *Nat. Mater.*, 17(7):633–641, 2018.

[66] L. F. Elliot. Modelling the proliferation of human stem cells. *MMath Report: Newcastle University*, 2015.

[67] B. de Lucas, L. M. Prez, and B. G. Glvez. Importance and regulation of adult stem cell migration. *J. Cell. Mol. Med.*, 22(2):746–754, 2018.

[68] E. Scarpa and R. Mayor. Collective cell migration in development. *J. Cell Biol.*, 212(2):143–155, 2016.

[69] K. Muguruma, A. Nishiyama, H. Kawakami, K. Hashimoto, and Y. Sasai. Self-organization of polarized cerebellar tissue in 3D culture of human pluripotent stem cells. *Cell Rep.*, 10(4):537–550, 2015.

[70] S. Dekoninck and C. Blanpain. Stem cell dynamics, migration and plasticity during wound healing. *Nat. Cell Biol.*, 21(1):18–24, 2019.

[71] S. Kusuma and S. Gerecht. Fast and furious: the mass and motion of stem cells. *Biophys. J.*, 105:837–838, 2013.

[72] L E Wadkin, S Orozco-Fuentes, I Neganova, M Lako, , N G Parker, and A Shukurov. *An introduction to the mathematical modelling of iPSCs*. Recent advances in iPSC technology. Elsevier (submitted), preprint: arXiv:2010.15493, 2020.

[73] A. Friedman. What is mathematical biology and how useful is it? *Notices Am. Math. Soc*, 57(7):851–857, 2010.

[74] J. D. Murray. *Mathematical Biology I. An Introduction*, volume 17 of *Interdisciplinary Applied Mathematics*. Springer, New York, 3 edition, 2002.

[75] J. D. Murray. *Mathematical Biology II. Spatial Models and Biomedical Applications*, volume 18 of *Interdisciplinary Applied Mathematics*. Springer, New York, 3 edition, 2003.

[76] R. Barnes and C. Lehman. Modeling of bovine spongiform encephalopathy in a two-species feedback loop. *Epidemics*, 5(2):85–91, 2013.

[77] M. Al-Zoughool, D. Cottrell, S. Elsaadany, N. Murray, T. Oraby, R. Smith, and D. Krewski. Mathematical models for estimating the risks of bovine spongiform encephalopathy (BSE). *J. Toxicol. Environ. Health B Crit. Rev.*, 18(2):71–104, 2015.

[78] S. M. Kissler, J. R. Gog, C. Viboud, V. Charu, O. N. Bjrnstad, L. Simonsen, and B. T. Grenfell. Geographic transmission hubs of the 2009 influenza pandemic in the United States. *Epidemics*, 26:86–94, 2019.

[79] M. R. Servedio, Y. Brandvain, S. Dhole, C. L. Fitzpatrick, E. E. Goldberg, C. A. Stern, J. Van Cleve, and J. D. Yeh. Not just a theory - the utility of mathematical models in evolutionary biology. *PLoS Biol.*, 12(12):1–5, 2014.

[80] T. Hillen and K. J. Painter. A user's guide to PDE models for chemotaxis. *J. Math. Biol.*, 58(1):183, 2008.

[81] P. M. Altrock, L. L. Liu, and F. Michor. The mathematics of cancer: integrating quantitative models. *Nat. Rev. Cancer*, 15(12):730–745, 2015.

[82] L. E. Wadkin, S. Orozco-Fuentes, I. Neganova, S. Bojic, A. Laude, M. Lako, N. G. Parker, and A. Shukurov. Seeding hESCs to achieve optimal colony clonality. *Sci. Rep.*, page 15299, 2019.

[83] J. Hanna, K. Saha, B. Pando, J. van Zon, C. J. Lengner, M. P. Creyghton, A. van Oudenaarden, and R. Jaenisch. Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature*, 462(7273):595–601, 2009.

[84] T. N. T. Nguyen, K. Sasaki, and M. Kino-oka. Elucidation of human induced pluripotent stem cell behaviors in colonies based on a kinetic model. *J. Biosci. Bioeng.*, 127(5):625–632, 2019.

[85] L. E. Wadkin, S. Orozco-Fuentes, I. Neganova, M. Lako, A. Shukurov, and N. G. Parker. The recent advances in the mathematical modelling of human pluripotent stem cells. *SN Applied Sciences*, 2(2):276, 2020.

[86] S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.*, 94:128701, 2005.

[87] L. Li, B. H. Wang, S. Wang, L. Moalim-Nour, K. Mohib, D. Lohnes, and L. Wang. Individual cell movement, asymmetric colony expansion, rho-associated kinase, and E-cadherin impact the clonogenicity of human embryonic stem cells. *Biophys. J.*, 98:2442–2451, 2010.

[88] L. E. Wadkin, L. F. Elliot, I. Neganova, N. G. Parker, V. Chichagova, G. Swan, A. Laude, M. Lako, and A. Shukurov. Dynamics of single human embryonic stem cells and their pairs: a quantitative analysis. *Sci. Rep.*, 7(1):1–12, 2017.

[89] L. E. Wadkin, S. Orozco-Fuentes, I. Neganova, G. Swan, A. Laude, M. Lako, A. Shukurov, and N. G. Parker. Correlated random walks of human embryonic stem cells in vitro. *Phys. Biol.*, 15(5):056006, 2018.

[90] M. Tewary, J. Ostblom, L. Prochazka, T. Zulueta-Coarasa, N. Shakiba, R. Fernandez-Gonzalez, and P. W. Zandstra. A stepwise model of reaction-diffusion and positional information governs self-organized human peri-gastrulation-like patterning. *Development*, 144(23):4298–4312, 2017.

[91] J. E. Till, E. A. McCulloch, and L. Siminovitch. A stochastic model of stem cell proliferation, based on the growth of spleen colony forming cells. *Proc. Natl. Acad. Sci. U.S.A.*, 51:29–36, 1964.

[92] P. Pir and N. Le Novere. Mathematical models of pluripotent stem cells: at the dawn of predictive regenerative medicine. *Methods Mol. Biol.*, 1386:331–350, 2016.

[93] V. Olariu and C. Peterson. Kinetic models of hematopoietic differentiation. *Wiley Interdiscip. Rev. Syst. Biol. Med.*, 11(1):e1424, 2019.

[94] H. Xu, Y. S. Ang, A. Sevilla, I. R. Lemischka, and A. Ma'ayan. Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells. *PLoS Comput. Biol.*, 10(8):1–14, 2014.

[95] R. Hannam, A. Annibale, and R. Khn. Cell reprogramming modelled as transitions in a hierarchy of cell cycles. *J. Phys. A.*, 50(42):425601, 2017.

[96] A. L. MacLean, C. Lo Celso, and M. P. H. Stumpf. Concise review: stem cell population biology: insights from hematopoiesis. *Stem Cells*, 35(1):80–88, 2017.

[97] S. Stepney and P. S. Andrews. Cosmos special issue editorial. *Nat. Comput.*, 14(1):1–6, 2015.

[98] J. M. Raimond, P. Goy, M. Gross, C. Fabre, and S. Haroche. Collective absorption of blackbody radiation by rydberg atoms in a cavity: An experiment on bose statistics and brownian motion. *Phys. Rev. Lett.*, 49:117–120, 1982.

[99] M. de Jager, F. Bartumeus, A. Klzsch, F. J. Weissing, G. M. Hengeveld, B. A. Nolet, P. M. J. Herman, and J. van de Koppel. How superdiffusion gets arrested: ecological encounters explain shift from levy to brownian movement. *Proc. Royal Soc. B*, 281(1774):20132605, 2014.

[100] M. Batty. *Agent-based pedestrian modelling.* Advanced spatial analysis: The CASA book of GIS. Esri Press Redlands, CA, USA, 2003.

[101] A. J. Loosley, A. M. O. Brien, J. S. Reichner, and J. X. Tang. Describing directional cell migration with a characteristic directionality time. *PLoS One*, 10, 2015.

[102] C. Gardiner. *Stochastic Methods: A Handbook for the Natural and Social Sciences.* Springer Series in Synergetics. Springer-Verlag Berlin Heidelberg, 2009.

[103] B. B. Mandelbrot and J. W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Rev.*, 10(4):422–437, 1968.

[104] S. Orozco-Fuentes, I. Neganova, L. E. Wadkin, A. W. Baggaley, R. A. Barrio, M. Lako, A. Shukurov, and N. G. Parker. Quantification of the morphological characteristics of hESC colonies. *Sci. Rep.*, 9(1):1–11, 2019.

[105] L. L. Liu, J. Brumbaugh, O. Bar-Nur, Z. Smith, M. Stadtfeld, A. Meissner, K. Hochedlinger, and F. Michor. Probabilistic modeling of reprogramming to induced pluripotent stem cells. *Cell Rep.*, 17(12):3395–3406, 2016.

[106] M. Tewary, N. Shakiba, and P. W. Zandstra. Stem cell bioengineering: building from stem cell biology. *Nat. Rev. Genet.*, 19(10):595–614, 2018.

[107] M. H. Swat, G. L. Thomas, J. M. Belmonte, A. Shirinifard, D. Hmeljak, and J. A. Glazier. Multi-scale modeling of tissues using CompuCell3D. In *Methods in cell biology*, volume 110, pages 325–366. Elsevier, 2012.

[108] A. Ghaffarizadeh, R. Heiland, S. H. Friedman, S. M. Mumenthaler, and P. Macklin. Physicell: An open source physics-based cell simulator for 3D multicellular systems. *PLoS Comput. Biol.*, 14(2):1–31, 2018.

[109] E. A. Codling, M. J. Plank, and S. Benhamou. Random walk models in biology. *J. R. Soc. Interface*, 5(25):813–834, 2008.

[110] P. Turchin. Quantitative analysis of movement: Measuring and modeling population redistribution in animals and plants. *Q. Rev. Biol.*, 74(2):240–241, 1999.

[111] P. Dieterich, R. Klages, R. Preuss, and A. Schwab. Anomalous dynamics of cell migration. *Proc. Natl. Acad. Sci. U.S.A.*, 105(2):459–463, 2008.

[112] P. Wu, A. Giri, S. X. Sun, and D. Wirtz. Three-dimensional cell migration does not follow a random walk. *Proc. Natl. Acad. Sci. U.S.A.*, 111(11):3949–3954, 2014.

[113] J-P. Bouchaud and A. Georges. Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.*, 195(4):127–293, 1990.

[114] W. M. Getz and D. Saltz. A framework for generating and analyzing movement paths on ecological landscapes. *Proc. Natl. Acad. Sci. U.S.A.*, 105(49):19066–19071, 2008.

[115] S. Benhamou. How to reliably estimate the tortuosity of an animals path: straightness, sinuosity, or fractal dimension? *J. Theor. Biol.*, 229:209–220, 2004.

[116] F. Bartumeus, M. G. E. da Luz, G. M. Viswanathan, and J. Catalan. Animal search strategies: a quantitative random-walk analysis. *Ecology*, 86(11):3078–3087, 2005.

[117] G. F. Lawler. *Random walk and the heat equation*, volume 55. American Mathematical Soc., 2010.

[118] R. B. Dickinson, S. Guido, and R. T. Tranquillo. Biased cell migration of fibroblasts exhibiting contact guidance in oriented collagen gels. *Ann. Biomed. Eng.*, 22(4):342–356, 1994.

[119] R. B. Dickinson. A generalized transport model for biased cell migration in an anisotropic environment. *J. Math. Biol.*, 40(2):97–135, 2000.

[120] W. Alt. Biased random walk models for chemotaxis and related diffusion approximations. *J. Math. Biol.*, 9(2):147–177, 1980.

[121] F. A. C. C. Chalub, P. A. Markowich, B. Perthame, and C. Schmeiser. Kinetic models for chemotaxis and their drift-diffusion limits. In *Nonlinear Differential Equation Models*, pages 123–141. Springer, 2004.

[122] A. Okubo and S. A. Levin. *Diffusion and ecological problems: modern perspectives*, volume 14. Springer Science & Business Media, 2013.

[123] M. H. Gail and C. W. Boone. The locomotion of mouse fibroblasts in tissue culture. *Biophys. J.*, 10(10):980, 1970.

[124] R. L. Hall. Amoeboid movement as a correlated walk. *J. Math. Biol.*, 4(4):327–335, 1977.

[125] A. A. Potdar, J. Jeon, A. M. Weaver, V. Quaranta, and P. T. Cummings. Human mammary epithelial cells exhibit a bimodal correlated random walk pattern. *PLoS One*, 5(3):1–10, 2010.

[126] P. J. M. Van Haastert. A model for a correlated random walk based on the ordered extension of pseudopodia. *PLoS Comput. Biol.*, 6(8):1–11, 2010.

[127] P. M. Kareiva and N. Shigesada. Analyzing insect movement as a correlated random walk. *Oecologia*, 56(2-3):234–238, 1983.

[128] C. M. Bergman, J. A. Schaefer, and S. N. Luttich. Caribou movement as a correlated random walk. *Oecologia*, 123(3):364–374, 2000.

[129] D. S. Johnson, J. M. London, M. Lea, and J. W. Durban. Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215, 2008.

[130] J. M. Drake and A. M. Kramer. Allee effects. *Nat. Edu. Knowledge*, 3(10):2, 2001.

[131] J. C. Gascoigne and R. N. Lipcius. Allee effects driven by predation. *J. Appl. Ecol.*, 41(5):801–810, 2004.

[132] K. E. Johnson, G. Howard, W. Mo, M. K. Strasser, E. A. B. F. Lima, S. Huang, and A. Brock. Cancer cell population growth kinetics at low densities deviate from the exponential growth model and suggest an Allee effect. *PLoS Biol.*, 17(8):1–29, 2019.

[133] F. Michor. Mathematical models of cancer stem cells. *J. Clin. Oncol.*, 26(17):2854–2861, 2008.

[134] M. Loeffler and H. E. Wichmann. A comprehensive mathematical model of stem cell proliferation which reproduces most of the published experimental results. *Cell Tissue Kinet.*, 13(5):543–561, 1980.

[135] B. M. Deasy, R. J. Jankowski, T. R. Payne, B. Cao, J. P. Goff, J. S. Greenberger, and J. Huard. Modeling stem cell population growth: incorporating terms for proliferative heterogeneity. *Stem Cells*, 21(5):536–545, 2003.

[136] J. L. Sherley, P. B. Stadler, and J. S. Stadler. A quantitative method for the analysis of mammalian cell proliferation in culture in terms of dividing and non-dividing cells. *Cell Prolif.*, 28(3):137–144, 1995.

[137] M. A. Tabatabai, Z. Bursac, W. M. Eby, and K. P. Singh. Mathematical modeling of stem cell proliferation. *Med. Biol. Eng. Comput.*, 49(3):253–262, 2011.

[138] M. Tabatabai, D. K. Williams, and Z. Bursac. Hyperbolastic growth models: theory and application. *Theor. Biol. Med. Model.*, 2:14, 2005.

[139] O. Symmons and A. Raj. Whats luck got to do with it: single cells, multiple fates, and biological non-determinism. *Mol. Cell*, 62(5):788–802, 2016.

[140] M. Herberg and I. Roeder. Computational modelling of embryonic stem-cell fate control. *Development*, 142(13):2250–2260, 2015.

[141] B. D. MacArthur, A. Ma'ayan, and I. R. Lemischka. Systems biology of stem cell fate and cellular reprogramming. *Nat. Rev. Mol. Cell Biol.*, 10(10):672–681, 2009.

[142] S. Viswanathan and P. W. Zandstra. Towards predictive models of stem cell fate. *Cytotechnology*, 41(2-3):75, 2003.

[143] Y. T. Lin, P. G. Hufton, E. J. Lee, and D. A. Potoyan. A stochastic and dynamical view of pluripotency in mouse embryonic stem cells. *PLoS Comput. Biol.*, 14(2):1–24, 2018.

[144] J. Wu and E. S. Tzanakakis. Contribution of stochastic partitioning at human embryonic stem cell division to NANOG heterogeneity. *PLoS One*, 7(11):1–14, 2012.

[145] J. Jang, D. Han, M. Golkaram, M. Audouard, G. Liu, D. Bridges, S. Hellander, A. Chialastri, S. S. Dey, L. R. Petzold, and K. S. Kosik. Control over single-cell distribution of G1 lengths by WNT governs pluripotency. *PLoS Biol.*, 17(9):1–24, 2019.

[146] Bin Zhang and Peter G. Wolynes. Stem cell differentiation as a many-body problem. *Proc. Natl. Acad. Sci. U.S.A.*, 111(28):10185–10190, 2014.

[147] V. Chickarmane, C. Troein, U. A. Nuber, H. M. Sauro, and C. Peterson. Transcriptional dynamics of the embryonic stem cell switch. *PLoS Comput. Biol.*, 2(9):1–13, 2006.

[148] V. Likhoshvai and A. Ratushny. Generalized Hill function method for modeling molecular processes. *J. Bioinf. Comput. Biol.*, 05(02b):521–531, 2007.

[149] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna. Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.*, 3(10):1–8, 2007.

[150] I. Glauche, M. Herberg, and I. Roeder. Nanog variability and pluripotency regulation of embryonic stem cells - insights from a mathematical model analysis. *PLoS One*, 5(6):1–12, 2010.

[151] T. Kalmar, C. Lim, P. Hayward, S. Muoz-Descalzo, J. Nichols, J. Garcia-Ojalvo, and A. Martinez Arias. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.*, 7(7):1–16, 2009.

[152] Y. Luo, C. L. Lim, J. Nichols, A. Martinez-Arias, and L. Wernisch. Cell signalling regulates dynamics of Nanog distribution in embryonic stem cell populations. *J. R. Soc. Interface*, 10(78):20120525, 2013.

[153] V. Olariu, C. Lövkvist, and K. Sneppen. Nanog, Oct4 and Tet1 interplay in establishing pluripotency. *Sci. Rep.*, 6:25438, 2016.

[154] R. C. G. Smith, P. S. Stumpf, S. J. Ridden, A. Sim, S. Filippi, H. A. Harrington, and B. D. MacArthur. Nanog fluctuations in embryonic stem cells highlight the problem of measurement in cell biology. *Biophys. J.*, 112(12):2641–2652, 2017.

[155] P. Yu, Q. Nie, C. Tang, and L. Zhang. Nanog induced intermediate state in regulating stem cell differentiation and reprogramming. *BMC Syst. Biol.*, 12(1):22, 2018.

[156] I. R. Akberdin, N. A. Omelyanchuk, S. I. Fadeev, N. E. Leskova, E. A. Oschepkova, F. V. Kazantsev, Y. G. Matushkin, D. A. Afonnikov, and N. A. Kolchanov. Pluripotency gene network dynamics: System views from parametric analysis. *PLoS One*, 13(3):1–24, 2018.

[157] D. Auddya and B. J. Roth. A mathematical description of a growing cell colony based on the mechanical bidomain model. *J. Phys. D. Appl. Phys.*, 50(10):105401, 2017.

[158] C. S. Henriquez. Simulating the electrical behavior of cardiac tissue using the bidomain model. *Crit. Rev. Biomed. Eng.*, 21(1):177, 1993.

[159] W. M. Eby and N. Coleman. *Mathematical Models in Stem Cell Differentiation and Fate Predictability*. Springer International Publishing, 2016.

[160] S. R. Vedula, A. Ravasio, C. T. Lim, and B. Ladoux. Collective cell migration: a mechanistic perspective. *Physiology*, 28(6):370–379, 2013.

[161] K. Rossington and T.Benson. An agent-based model to predict fish collisions with tidal stream turbines. *Renew. Energy*, 151:1220–1229, 2020.

[162] C. M. Henein and T. White. Agent-based modelling of forces in crowds. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 173–184. Springer, 2004.

[163] D. C. Walker, J. Southgate, G. Hill, M. Holcombe, D. R. Hose, S. M. Wood, S. MacNeil, and R. H. Smallwood. The epitheliome: agent-based modelling of the social behavior of cells. *Biosystems*, 76:89–100, 2004.

[164] M. dInverno and R. Saunders. Agent-based modelling of stem cell self-organisation in a niche. In *International Workshop on Engineering Self-Organising Applications*, pages 52–68. Springer, 2004.

[165] J. Poleszczuk, P. Macklin, and H. Enderling. Agent-based modeling of cancer stem cell driven solid tumor growth. *Methods Mol. Biol.*, 1516:335–346, 2016.

[166] M. Hoffmann, J. Kuska, M. Zscharnack, M. Loeffler, and J. Galle. Spatial organization of mesenchymal stem cells in-vitro results from a new individual cell-based model with podia. *PLoS One*, 6(7):1–16, 2011.

[167] T. A. Zangle, J. Chun, J. Zhang, J. Reed, and M. A. Teitell. Quantification of biomass and cell motion in human pluripotent stem cell colonies. *Biophys. J.*, 105(3):593–601, 2013.

[168] M. Kino-Oka, R. Umegaki, M. Taya, S. Tone, and J. E. Prenosil. Valuation of growth parameters in monolayer keratinocyte cultures based on a two-dimensional cell placement model. *J. Biosci. Bioeng.*, 89(3):285–287, 2000.

[169] Y. Lei, D. Jeong, J. Xiao, and D. V. Schaffer. Developing defined and scalable 3D culture systems for culturing human pluripotent stem cells at high densities. *Cell. Mol. Bioeng.*, 7(2):172–183, 2014.

[170] X. Yin, B. E. Mead, H. Safaee, R. Langer, J. M. Karp, and O. Levy. Engineering stem cell organoids. *Cell Stem Cell*, 18(1):25–38, 2016.

[171] S. Montes-Olivas, L. Marucci, and M. Homer. Mathematical models of organoid cultures. *Front. Genet.*, 10:873, 2019.

[172] H. Khayyeri, S. Checa, M. Tgil, and P. J. Prendergast. Corroboration of mechanobiological simulations of tissue differentiation in an in-vivo bone chamber using a lattice-modeling approach. *J. Orthop. Res.*, 27(12):1659–1666, 2009.

[173] F. Graner and J. A. Glazier. Simulation of biological cell sorting using a two-dimensional extended Potts model. *Phys. Rev. Lett.*, 69(13):2013, 1992.

[174] A. Szab and R. M. Merks. Cellular Potts modeling of tumor growth, tumor invasion, and tumor evolution. *Front. Oncol.*, 3:87, 2013.

[175] A. R.G. Libby, D. Briers, I. Haghighi, D. A. Joy, B. R. Conklin, C. Belta, and T. C. McDevitt. Automated design of pluripotent stem cell self-organization. *Cell Syst.*, 9(5):483–495.e10, 2019.

[176] S. Adra, T. Sun, S. MacNeil, M. Holcombe, and R. Smallwood. Development of a three dimensional multiscale computational model of the human epidermis. *PLoS One*, 5(1):1–13, 2010.

[177] P. Van Liedekerke, A. Buttenschön, and D. Drasdo. Off-lattice agent-based models for cell and tumor growth: Numerical methods, implementation, and applications. In *Numerical Methods and Advanced Simulation in Biomechanics and Biological Processes*, pages 245–267. Academic Press, 2018.

[178] P. Van Liedekerke, M. M. Palm, N. Jagiella, and D. Drasdo. Simulating tissue mechanics with agent-based models: concepts, perspectives and some novel results. *Comput. Part. Mech.*, 2(4):401–444, 2015.

[179] R. Ganguly and I. K. Puri. Mathematical model for the cancer stem cell hypothesis. *Cell Prolif.*, 39(1):3–14, 2006.

[180] S. Orozco-Fuentes and R. A. Barrio. Modelling the dynamics of stem cells in colonic crypts. *Eur. Phys. J-Spec. Top.*, 226(3):353–363, 2017.

[181] F. A. Meineke, C. S. Potten, and M. Loeffler. Cell migration and organization in the intestinal crypt using a lattice-free model. *Cell Prolif.*, 34(4):253–266, 2001.

[182] R. A. Barrio, J. R. Romero-Arias, M. A. Noguez, E. Azpeitia, E. Ortiz-Gutiérrez, V. Hernández-Hernández, Y. Cortes-Poza, and E. R. Álvarez-Buylla. Cell patterns emerge from coupled chemical and physical fields with cell proliferation dynamics: The Arabidopsis thaliana root as a study system. *PLoS Comput. Biol.*, 9(5):1–12, 2013.

[183] D. Lehotzky and G. K. H. Zupanc. Cellular automata modeling of stem-cell-driven development of tissue in the nervous system. *Dev. Neurobiol.*, 79(5):497–517, 2019.

[184] M. A. Pérez and P. J. Prendergast. Random-walk models of cell dispersal included in mechanobiological simulations of tissue differentiation. *J. Biomech.*, 40(10):2244–2253, 2007.

[185] Z. Wang, J. D. Butner, R. Kerketta, V. Cristini, and T. S. Deisboeck. Simulating cancer growth with multiscale agent-based modeling. *Semin. Cancer Biol.*, 30:70–78, 2015.

[186] D. Tartarini and E. Mele. Adult stem cell therapies for wound healing: Biomaterials and computational models. *Front. Bioeng. Biotechnol.*, 3:206, 2016.

[187] L. E Wadkin, S. Orozco-Fuentes, I. Neganova, M. Lako, N. G. Parker, and A. Shukurov. OCT4 expression in human embryonic stem cells: spatio-temporal dynamics and fate transitions. *arXiv: q-bio.CB 2004.09133*, 2020.

[188] S. M. Phadnis, N. O. Loewke, I. K. Dimov, S. Pai, C. E. Amwake, O. Solgaard, T. M. Baer, B. Chen, and R. A. Reijo Pera. Dynamic and social behaviours of human pluripotent stem cells. *Sci. Rep.*, 5:14209, 2015.

[189] K. F. Jarrell and M. J. McBride. The surprisingly diverse ways that prokaryotes move. *Nat. Rev. Microbiol.*, 6(6):466–476, 2008.

[190] G. Danuser, J. Allard, and A. Mogilner. Mathematical modeling of eukaryotic cell migration: Insights beyond experiments. *Annu. Rev. Cell Dev. Bi.*, 29(1):501–528, 2013.

[191] S. Kurosaka and A. Kashina. Cell biology of embryonic migration. *Birth Defects Res. C Embryo Today*, 84(2):102–122, 2008.

[192] A. Aman and T. Piotrowski. Cell migration during morphogenesis. *Dev. Biol.*, 341(1):20–33, 2010.

[193] C. M. Franz, G. E. Jones, and A. J. Ridley. Cell migration in development and disease. *Dev. Cell*, 2(2):153–158, 2002.

[194] L. Li, Y. He, M. Zhao, and J. Jiang. Collective cell migration: implications for wound healing and cancer invasion. *Burns Trauma*, 1(1):21–26, 2013.

[195] L. Lamalice, F. Le Boeuf, and J. Huot. Endothelial cell migration during angiogenesis. *Circ. Res.*, 100(6):782–794, 2007.

[196] R. Guerrini and E. Parrini. Neuronal migration disorders. *Neurobiol. Dis.*, 38(2):154–166, 2010.

[197] P. Friedl and K. Wolf. Tumour-cell invasion and migration: diversity and escape mechanisms. *Nat. Rev. Cancer*, 3(5):362–374, 2003.

[198] F. Friedl, J. Locker, E. Sahai, and J. E. Segall. Classifying collective cancer cell invasion. *Nat. Cell Biol.*, 14(8):777–783, 2012.

[199] D. T. Scadden. The stem-cell niche as an entity of action. *Nature*, 441(7097):1075–1079, 2006.

[200] X. Guan. Cancer metastases: challenges and opportunities. *Acta. Pharm. Sin. B*, 5(5):402–418, 2015.

[201] C. D. Paul, P. Mistriotis, and K. Konstantopoulos. Cancer cell motility: lessons from migration in confined spaces. *Nat. Rev. Cancer*, 17(2):131–140, 2017.

[202] H. Kress, E. H. K. Stelzer, D. Holzer, F. Buss, G. Griffiths, and A. Rohrbach. Filopodia act as phagocytic tentacles and pull with discrete steps and a load-dependent velocity. *Proc. Natl. Acad. Sci. U.S.A.*, 104, 2007.

[203] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell. 4th edition.*, volume The Cytoskeleton and Cell Behavior. New York: Garland Science, 2002.

[204] L. Bosgraaf and P. J. M. Van Haastert. The ordered extension of pseudopodia by amoeboid cells in the absence of external cues. *PLoS One*, 4(4):1–13, 2009.

[205] J. H. Zar. *Biostatistical analysis*. Pearson Education India, 1999.

[206] P. Berens. A matlab toolbox for circular statistics. *Journal of Statistical Software*, 31, 2009.

[207] M. Krause and A. Gautreau. Steering cell migration: lamellipodium dynamics and the regulations of directional persistence. *Nat. Rev. Mol. Cell Biol.*, 15:577–590, 2014.

[208] G. S. Watson and E. J. Williams. On the construction of significance tests on the circle and the sphere. *Biometrika*, 43(3-4):344–352, 1956.

[209] R. J. Petrie, A. D. Doyle, and K. M. Yamada. Random versus directionally persistent cell migration. *Nat. Rev. Mol. Cell Biol.*, 10:538–549, 2009.

[210] G. Reig, E. Pulgar, and M. L. Concha. Cell migration: from tissue culture to embryos. *Development*, 141(10):1999–2013, 2014.

[211] J. P. Rieu, A. Upadhyaya, J. A. Glazier, N. B. Ouchi, and Y. Sawada. Diffusion and deformations of single hydra cells in cellular aggregates. *Biophys. J.*, 79(4):1903–1914, 2000.

[212] S. Burov, S. M. Tabei, T. Huynh, M. P. Murrell, L. H. Philipson, S. A. Rice, M. L. Gardel, N. F. Scherer, and A. R. Dinner. Distribution of directional change as a signature of complex dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 110(49):19689–19694, 2013.

[213] K. J. Duffy and R. M. Ford. Turn angle and run time distributions characterize swimming behavior for Pseudomonas putida. *J. Bacteriol.*, 179(4):1428–1430, 1997.

[214] R. L. Hall and S. C. Peterson. Trajectories of human granulocytes. *Biophys. J.*, 25(2):365–372, 1979.

[215] J. E. Johansen, J. Pinhassi, N. Blackburn, U. L. Zweifel, and Å. Hagström. Variability in motility characteristics among marine bacteria. *Aquatic Microb. Ecol.*, 28(3):229–237, 2002.

[216] G. M. Barbara and J. G. Mitchell. Bacterial tracking of motile algae. *FEMS Microbiol. Ecol.*, 44(1):79–87, 2003.

[217] F. Detcheverry. Generalized run-and-turn motions: From bacteria to lévy walks. *Phys. Rev. E*, 96(1):012415, 2017.

[218] A. Acharyya. Random walk with non-uniform angular distribution biased by an external periodic pulse. *Eur. J. Phys*, 37(6), 2016.

[219] P. Maiuri, J. Rupprecht, S. Wieser, V. Ruprecht, O. Bnichou, N. Carpi, M. Coppey, S. De Beco, N. Gov, C. Heisenberg, C. Lage Crespo, F. Lautenschlaeger, M. Le Berre, A. Lennon-Dumenil, M. Raab, H. Thiam, M. Piel, M. Sixt, and R. Voituriez. Actin flows mediate a universal coupling between cell speed and cell persistence. *Cell*, 161(2):374–386, 2015.

[220] Wu D. Signaling mechanisms for regulation of chemotaxis. *Cell Res.*, 15(1):52–56, 2005.

[221] H. Lodish, A. Berk, L. S. Zipursky, P. Matsudaira, D. Baltimore, and J. Darnell. *Cell-cell adhesion and communication.* WH Freeman, 2000.

[222] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell. 4th edition.*, volume Cell-Cell Adhesion. New York: Garland Science, 2002.

[223] Y. Arboleda-Estudillo, M. Krieg, J. Stühmer, N. A. Licata, D. J. Muller, and C. Heisenberg. Movement directionality in collective migration of germ layer progenitors. *Curr. Biol.*, 20(2):161–169, 2010.

[224] P. Rørth. Collective cell migration. *Annu. Rev. Cell Dev. Bi.*, 25(1):407–429, 2009. PMID: 19575657.

[225] K. Suzuki, J. Saito, R. Yanai, N. Yamada, T. Chikama, K. Seki, and T. Nishida. Cell-matrix and cell-cell interactions during corneal epithelial wound healing. *Prog. Retin. Eye Res.*, 22(2):113–133, 2003.

[226] P. Friedl and D. Gilmour. Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.*, 10(7):445, 2009.

[227] S. Vedel, S. Tay, D. M. Johnston, H. Bruus, and S. R. Quake. Migration of cells in a social context. *Proc. Natl. Acad. Sci. U.S.A.*, 110(1):129–134, 2013.

[228] E. Fuchs, T. Tumbar, and G. Guasch. Socializing with the neighbors: Stem cells and their niche. *Cell*, 116(6):769–778, 2004.

[229] M. Amit, M. K. Carpenter, M. S. Inokuma, C. Chiu, C. P. Harris, M. A. Waknitz, J. Itskovitz-Eldor, and J. A. Thomson. Clonally derived human embryonic stem cell lines maintain pluripotency and proliferative potential for prolonged periods of culture. *Dev. Biol.*, 227(2):271–278, 2000.

[230] E. Matsa, J. H. Ahrens, and J. C. Wu. Human induced pluripotent stem cells as a platform for personalized and precision cardiovascular medicine. *Physiol. Rev.*, 96(3):1093–1126, 2016.

[231] B. K. Gage, T. D. Webber, and T. J. Kieffer. Initial cell seeding density influences pancreatic endocrine development during in vitro differentiation of human embryonic stem cells. *PLoS One*, 8(12):e82076, 2013.

[232] K. Jacobs, F. Zambelli, A. Mertzanidou, I. Smolders, M. Geens, H. T. Nguyen, L. Barbe, K. Sermon, and C. Spits. Higher-density culture in human embryonic stem cells results in DNA damage and genome instability. *Stem Cell Rep.*, 6(3):330–341, 2016.

[233] P. J. Gokhale, J. K. Au-Young, S. Dadi, D. N. Keys, N. J. Harrison, M.k Jones, S. Soneji, T. Enver, J. K. Sherlock, and P. W. Andrews. Culture adaptation

alters transcriptional hierarchies among single human embryonic stem cells reflecting altered patterns of differentiation. *PLoS One*, 10(4):1–13, 2015.

[234] P. H. Schwartz, D. J. Brick, H. E. Nethercott, and A. E. Stover. Traditional human embryonic stem cell culture. *Methods Mol. Biol.*, 767:107–123, 2011.

[235] D. Moogk, M. Stewart, D. Gamble, M. Bhatia, and E. Jervis. Human ESC colony formation is dependent on interplay between self-renewing hESCs and unique precursors responsible for niche generation. *Cytometry A*, 77A(4):321–327, 2010.

[236] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical analysis and modelling of spatial point patterns*, volume 70. John Wiley & Sons, 2008.

[237] Z. Zhu and D. Huangfu. Human pluripotent stem cells: an emerging model in developmental biology. *Development*, 140(4):705–717, 2013.

[238] D. Ilic and C. Ogilvie. Concise Review: Human embryonic stem cells - What have we done? What are we doing? Where are we going? *Stem Cells*, 35(1):17–25, 2017.

[239] A. Trounson and N. D. DeWitt. Pluripotent stem cells progressing to the clinic. *Nat. Rev. Mol. Cell Biol.*, 17(3):194, 2016.

[240] M-E Torres-Padilla and I. Chambers. Transcription factor heterogeneity in pluripotent stem cells: a stochastic advantage. *Development*, 141(11):2173–2181, 2014.

[241] R. Stadhouders, G. J. Filion, and T. Graf. Transcription factors and 3D genome conformation in cell-fate decisions. *Nature*, 569(7756):345–354, 2019.

[242] M. Li and J. C. Izpisua Belmonte. Deconstructing the pluripotency gene regulatory network. *Nat. Cell Biol.*, 20(4):382–392, 2018.

[243] L. A. Boyer, T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch, and R. A. Young. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell*, 122(6):947956, 2005.

[244] I. Chambers and S. R. Tomlinson. The transcriptional foundation of pluripotency. *Development*, 136(14):2311–2322, 2009.

[245] K. Kee, J. M. Gonsalves, A. T. Clark, and R. A. R. Pera. Bone morphogenetic proteins induce germ cell differentiation from human embryonic stem cells. *Stem Cells Dev.*, 15(6):831–837, 2006.

[246] R. Xu, X. Chen, D. S. Li, R. Li, G. C. Addicks, C. Glennon, T. P. Zwaka, and J. A. Thomson. BMP4 initiates human embryonic stem cell differentiation to trophoblast. *Nat. Biotechnol.*, 20(12):1261–1264, 2002.

[247] A. M. Singh, J. Chappell, R. Trost, L. Lin, T. Wang, J. Tang, H. Wu, S. Zhao, P. Jin, and S. Dalton. Cell-cycle control of developmentally regulated transcription factors accounts for heterogeneity in human pluripotent cells. *Stem Cell Rep.*, 1(6):532–544, 2013.

[248] Z-N. Zhang, S-K. Chung, Z. Xu, and Y. Xu. OCT4 maintains the pluripotency of human embryonic stem cells by inactivating p53 through Sirt1-mediated deacetylation. *Stem Cells*, 32(1):157–165, 2014.

[249] R. T. Rodriguez, J. M. Velkey, C. Lutzko, R. Seerke, D. B. Kohn, K. S. OShea, and M. T. Firpo. Manipulation of OCT4 levels in human embryonic stem cells results in induction of differential cell types. *Exp. Biol. Med.*, 232(10):1368–1380, 2007.

[250] G. Shi and J. Ying. Role of OCT4 in maintaining and regaining stem cell pluripotency. *Stem Cell Res. Ther.*, 1(5):39, 2010.

[251] G. Pan and J. A. Thomson. Nanog and transcriptional networks in embryonic stem cell pluripotency. *Cell Res.*, 17:42–49, 2007.

[252] Q. Lin. Allee effect increasing the final density of the species subject to the Allee effect in a Lotka–Volterra commensal symbiosis model. *Adv. Differ. Equ.*, 196(1):1–9, 2018.

[253] S. Huang. Non-genetic heterogeneity of cells in development: more than just noise. *Development*, 136(23):3853–3862, 2009.

[254] M. H. Stewart, M. Bossé, K. Chadwick, P. Menendez, S. C. Bendall, and M. Bhatia. Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. *Nat. Methods*, 3(10):807–815, 2006.

[255] S. Dakhore, B. Nayer, and K. Hasegawa. Human pluripotent stem cell culture: Current status, challenges, and advancement. *Stem Cells Int.*, 2018:7396905–7396905, 2018.

[256] D. Stockholm, R. Benchaouir, J. Picot, P. Rameau, T. M. A. Neildez, G. Landini, C. Laplace-Builh, and A. Paldi. The origin of phenotypic heterogeneity in a clonal cell population in vitro. *PLoS One*, 2(4):1–13, 2007.

[257] W. Li, L-B. Cui, and M. K. Ng. On computation of the steady-state probability distribution of probabilistic Boolean networks with gene perturbation. *J. Comput. Appl. Math.*, 236(16):4067–4081, 2012.

[258] H. Niwa, J. Miyazaki, and A. G. Smith. Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nat. Genet.*, 24(4):372–376, 2000.

[259] J. L. Kopp, B. D. Ormsbee, M. Desler, and A. Rizzino. Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem cells*, 26(4):903–911, 2008.

[260] D. Strebinger, C. Deluz, E. T. Friman, S. Govindan, A. B. Alber, and D. M. Suter. Endogenous fluctuations of OCT4 and SOX2 bias pluripotent cell fate decisions. *Mol. Syst. Biol.*, 15(9):e9002, 2019.

[261] M. Skamagki, K. B. Wicher, A. J., S. Ganguly, and M. Zernicka-Goetz. Asymmetric localization of CDX2 mRNA during the first cell-fate decision in early mouse development. *Cell Reports*, 3(2):442–457, 2013.

[262] W-W. Tee and D. Reinberg. Chromatin features and the epigenetic regulation of pluripotency states in ESCs. *Development*, 141(12):2376–2390, 2014.

[263] Y. Babaie, R. Herwig, B. Greber, T. C. Brink, W. Wruck, D. Groth, H. Lehrach, T. Burdon, and J. Adjaye. Analysis of Oct4-dependent transcriptional networks regulating self-renewal and pluripotency in human embryonic stem cells. *Stem cells*, 25(2):500–510, 2007.

[264] A. Burykin, M. D. Costa, L. Citi, and A. L. Goldberger. Dynamical density delay maps: simple, new method for visualising the behaviour of complex systems. *BMC Med. Inform. Decis. Mak.*, 14(1):6, 2014.

[265] M. Fishman, F. J. Jacono, S. Park, R. Jamasebi, A. Thungtong, K. A. Loparo, and T. E. Dick. A method for analyzing temporal patterns of variability of a time series from Poincaré plots. *J. Appl. Physiol.*, 113(2):297–306, 2012.

[266] J. Mielniczuk and P. Wojdyłło. Estimation of Hurst exponent revisited. *Comput. Stat. Data Anal*, 51(9):4510–4525, 2007.

[267] T. Aste. Matlab File Exchange: Generalized Hurst exponent. https://uk.mathworks.com/matlabcentral/fileexchange/30076-generalized-hurst-exponent. Accessed: September 2019.

[268] L. Lacasa, B. Luque, J. Luque, and J. C. Nuno. The visibility graph: A new method for estimating the Hurst exponent of fractional Brownian motion. *EPL*, 86(3):30001, 2009.

[269] J. Barunik and L. Kristoufek. On Hurst exponent estimation under heavy-tailed distributions. *Physica A*, 389(18):3844–3855, 2010.

[270] K. S. Zaret. Genome reactivation after the silence in mitosis: recapitulating mechanisms of development? *Dev. Cell*, 29(2):132–134, 2014.

[271] N. Festuccia, I. Gonzalez, N. Owens, and P. Navarro. Mitotic bookmarking in development and stem cells. *Development*, 144(20):3633–3645, 2017.

[272] A. A. Sveshnikov, I. N. Sneddon, and M. Stark. *Applied Methods of the Theory of Random Functions*. ISSN. Elsevier Science, 1966.

[273] J. Shin, T. W. Kim, H. Kim, H. J. Kim, M. Y. Suh, S. Lee, H. Lee, S. Kwak, S. Lee, J. Lee, H. Jang, E. Cho, and H. Youn. Aurkb/PP1-mediated resetting of Oct4 during the cell cycle determines the identity of embryonic stem cells. *eLife*, 5:e10877, 2016.

[274] H. J. Kim, J. Shin, S. Lee, T. W. Kim, H. Jang, M. Y. Suh, J. Kim, I. Hwang, D. S. Hwang, E. Cho, and H. Youn. Cyclin-dependent kinase 1 activity coordinates the chromatin associated state of Oct4 during cell cycle in embryonic stem cells. *Nucleic Acids Res.*, 46(13):6544–6560, 2018.

[275] W. R. Holmes, N. S. Reyes de Mochel, Q. Wang, H. Du, T. Peng, M. Chiang, O. Cinquin, K. Cho, and Q. Nie. Gene expression noise enhances robust organization of the early mammalian blastocyst. *PLoS Comput. Biol.*, 13(1):1–23, 2017.

[276] E. Purdom and S. P. Holmes. Error distribution for gene expression data. *Stat. Appl. Genet. Mol. Biol.*, 4(1), 2005.

[277] M. Ghorbani, E. A. Jonckheere, and P. Bogdan. Gene expression is not random: Scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks. *Front. Physiol.*, 9:1446, 2018.

[278] X. Liu, B. Wang, and L. Xu. Statistical analysis of Hurst exponents of essential/nonessential genes in 33 bacterial genomes. *PLoS One*, 10(6):1–9, 2015.

[279] P. Bogdan, B. M. Deasy, B. Gharaibeh, T. Roehrs, and R. Marculescu. Heterogeneous structure of stem cells dynamics: statistical models and quantitative predictions. *Sci. Rep.*, 4:4826, 2014.

[280] H. B. Sieburg, G. Cattarossi, and C. E. Muller-Sieburg. Lifespan differences in hematopoietic stem cells are due to imperfect repair and unstable mean-reversion. *PLoS Comput. Biol.*, 9(4):1–15, 2013.

[281] L. Haghverdi, F. Buettner, and F. J. Theis. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics*, 31(18):2989–2998, 2015.

[282] Antonius Bernardus Dieker and M Mandjes. On spectral simulation of fractional Brownian motion. *Probab. Eng. Inform. Sc.*, 17(3):417–434, 2003.

[283] T. Dieker. *Simulation of fractional Brownian motion*. PhD thesis, Masters Thesis, Department of Mathematical Sciences, University of Twente , 2004.

[284] Z. M. Yin. New methods for simulation of fractional Brownian motion. *J. Comput. Phys.*, 127(1):66–72, 1996.

[285] S. Stoev. Matlab File Exchange: Simulation of Fractional Gaussian Noise *EXACT*. https://www.mathworks.com/matlabcentral/fileexchange/19797-simulation-of-fractional-gaussian-noise-exact. Accessed: January 2020.

[286] C. R. Dietrich and G. N. Newsam. Fast and exact simulation of stationary gaussian processes through circulant embedding of the covariance matrix. *SIAM J. Sci. Comput.*, 18(4):1088–1107, 1997.

[287] S. B. Lowen. Efficent generation of fractional Brownian motion for simulation of infrared focal-plane array calibration drift. *Methodol. Comput. Appl.*, 1(4):445–456, 1999.

[288] P. Abry and F. Sellan. The wavelet-based synthesis for fractional Brownian motion proposed by F. Sellan and Y. Meyer: Remarks and fast implementation, 1996.

[289] S. N. Elaydi and R. J. Sacker. Population models with Allee effect: a new model. *J. Biol. Dynam.*, 4(4):397–408, 2010.

[290] M. Wang, M. Kot, and M. G. Neubert. Integrodifference equations, Allee effects, and invasions. *J. Math. Biol.*, 44(2):150–168, 2002.

[291] S-R. Zhou, Y-F. Liu, and G. Wang. The stability of predatorprey systems subject to the Allee effects. *Theor. Popul. Biol.*, 67(1):23–31, 2005.

[292] P. Cheridito. Arbitrage in fractional Brownian motion models. *Finance Stoch.*, 7(4):533–553, 2003.

[293] W-L. Xiao, W-G. Zhang, X-L. Zhang, and Y-L. Wang. Pricing currency options in a fractional Brownian motion with jumps. *Econ. Model.*, 27(5):935–942, 2010.

[294] C. Bender, T. Sottinen, and E. Valkeila. Fractional processes as models in stochastic finance. In *Advanced mathematical methods for finance*, pages 75–103. Springer, 2011.

[295] D. Ernst, M. Hellmann, J. Köhler, and M. Weiss. Fractional Brownian motion in crowded fluids. *Soft Matter*, 8(18):4886–4889, 2012.

[296] B. B. Mandelbrot. The fractal geometry of nature. *Nature*, pages 394–397, 1982.

[297] I. Zachevsky and Y. Y. Zeevi. Single-image super-resolution of natural stochastic textures based on fractional Brownian motion. *IEEE T. Image Proc.*, 23(5):2096–2108, 2014.

[298] WJ Guo, YX Wang, MX Xie, and YJ Cui. Modeling oil spill trajectory in coastal waters based on fractional Brownian motion. *Mar. Pollut. Bull.*, 58(9):1339–1346, 2009.

[299] S. A. Sevier, D. A. Kessler, and H. Levine. Mechanical bounds to transcriptional noise. *Proc. Natl. Acad. Sci. U.S.A.*, 113(49):13983–13988, 2016.

[300] M. N. Artyomov, A. Meissner, and A. K. Chakraborty. A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. *PLoS Comput. Biol.*, 6(5):1–14, 2010.

[301] R. Morris, I. Sancho-Martinez, T. O. Sharpee, and J. C. Izpisua Belmonte. Mathematical approaches to modeling development and reprogramming. *Proc. Natl. Acad. Sci. U.S.A.*, 111(14):5076–5082, 2014.

[302] E. Meijering. *Methods for Cell and Particle Tracking, in Imaging and Spectroscopic Analysis of Living Cells*, volume 504. Elsevier, 2010.

[303] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. Fiji: an open-source platform for biological-image analysis. *Nat. Methods*, 9(7):676–682, 2012.

[304] C. T. Rueden, J. Schindelin, M. C. Hiner, B. E. DeZonia, A. E. Walter, E. T. Arena, and K. W. Eliceiri. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*, 18(1):529, 2017.

[305] E. Batschelet. *Circular Statistics in Biology.* Academic Press, 1981.

[306] M. Neuhäuser A. Pewsey and G. D. Ruxton. *Circular statistics in R.* OUP Oxford, 2013.